Understanding malignant hyperthermia: Bioinformatic approaches to identify pathogenic genetic variants

A thesis presented in partial fulfilment of the requirements for the degree of
Master of Science
in
Biochemistry
at Massey University, Palmerston North,
New Zealand.

Shannon Ormond

2020

Acknowledgements

Firstly, I would of course like to thank my supervisor Kathryn. She has always been so friendly, approachable, and patient with me. She has inspired me more than she could know.

Thanks to the friendly Massey staff who have helped me do all sorts of things, but especially the administration team, Ann and Cynthia.

I am grateful for my co-supervisor, Professor Murray Cox, who helped me develop a passion for programming!

An enormous thank you to Dr. David Winter, my unofficial third co-supervisor, who taught me an innumerable number of computer 'hacks', and instilled in me some amount of confidence with the command line.

Thank you to my adoptive lab Mothers, Anja and Rowane, who have not only been so willing to help me in the lab, but have also provided me with much needed therapy at times. And to my labmates: Sophie, Jeremy and Liam (and Ruby too!) and everybody else in our little corner of the science towers. I will miss the lab and everybody in it tremendously. More formally, thank you Sophie for supplying me with one of your wonderful pieces of RyR1 artwork.

And thanks to all my other friends in the science towers, including the lab groups who have accepted me as part of their own: X-lab and the CRG. A special mention to Del, for hosting the best Friday movie nights, and Raoul, because you're awesome.

Thank you for the close friendship Chelsea and Travis.

My family, thank you for always believing in me.

Last, but certainly not least, thank you Sean, my number one fan through all of my studies.

Abstract

Technological advances and decreasing costs in genome sequencing have greatly sped up the rate of identification of the genetic causes of inherited disease. One such human disorder in which genome sequencing is being applied in hope that the genetic causes will be identified is malignant hyperthermia (MH). MH is an autosomal dominant pharmacogenetic disorder which has long perplexed researchers due to its phenotypic and genetic complexity. Individuals susceptible to MH are at risk of a sometimes deadly hypermetabolic episode of skeletal muscle triggered by potent volatile general anaesthetic agents. Although the genetic origin of the disorder has been determined in over half of all MH families, there are many families for which the origin has not yet been elucidated. This research aimed to identify genetic variants that may be pathogenic for MH-susceptibility in six New Zealand families for which the genetic cause has not yet been identified. Targeted next-generation sequencing of the genome was undertaken on a number of individuals from each family. Bioinformatic approaches were developed and applied to identify candidate genetic variants. Segregation analysis was carried out for some of the identified candidate variants, which failed to establish an association with MH-susceptibility, although a number of variants were ruled out as being pathogenic for MH-susceptibility. Additionally, a common polymorphism that has been previously postulated to have a modifying effect on MH-susceptibility was identified within a large MH family. A genotype/phenotype association study was carried, however the study did not find an association. Overall, this work has reinforced that MH does not have a simple, easily identifiable genetic origin, suggesting there is one or more missing elements to the current understanding of MH. Importantly, it has established an effective MH-specific bioinformatic protocol that can be applied to other sequencing data in the future.

For my grandparents...

Abbreviations

μ	micro
μM	micromol L^{-1}
AF	allele frequency
ATP	adenosine triphosphate
bp	base pair
CCD	central core disease
CNV	copy-number variation/variant
DHPR	dihydropyridine receptor
DM1	myotonic dystrophy type I
DNA	deoxyribonucleic acid
ECC	excitation-contraction coupling
EDTA	ethylenediaminetetraacetic acid
EMHG	European Malignant Hyperthermia Group
FRET	Förster resonance energy transfer
g	grams
GATK	Genome Analysis Toolkit
gDNA	genomic DNA
HRM	high resolution melting (analysis)

indel insertion/deletion

IVCT *in vitro* contracture test

L litre

M mol L^{-1}

MDMA 3,4-Methylenedioxymethamphetamine

mg milligrams

MH malignant hyperthermia

MHN malignant hyperthermia negative

MHS malignant hyperthermia susceptible

MHS_c malignant hyperthermia equivocal to caffeine

MHS_h malignant hyperthermia equivocal to halothane

mM millimol L^{-1}

NGS next-generation sequencing

PCR polymerase chain reaction

PM plasma membrane

PTM post-translation modification

redox reduction-oxidation

RNA ribonucleic acid

RyR1 ryanodine receptor 1

SM skeletal muscle

SNP single nucleotide polymorphism

SNV single nucleotide variant

SR sarcoplasmic reticulum

T_m primer temperature of melting

TE buffer tris-EDTA buffer

TRPC transient receptor potential cation

VA volatile anaesthetics

VCF variant call format

VOI variant/s of interest

w/v weight/volume

WES whole exome sequencing/sequence/s

WGS whole genome sequencing

wt wild-type

x g multiple of earths gravitational force

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Muscle physiology

Muscle is a soft tissue of animals which functions to produce motion and force. There are three major muscle types, two being forms of striated muscle (skeletal and cardiac muscle) and the third being non-striated (smooth muscle). Striated muscle has repeat units called sarcomeres, which are not present in non-striated muscle. Biochemical differences of the three muscle types give rise to their specialised functions; going forward the focus will be on skeletal muscle (SM).

SM is so named as it is attached to the skeletal system. Contraction events of SM are controlled in a voluntary fashion by the somatic nervous system, as opposed to non-voluntary contraction events of cardiac and smooth muscle. A contraction event occurs when an electrochemical current is transduced from the brain to a motor neuron through the nervous system. Motor neurons contact cells of SM, forming a neuromuscular junction, ultimately allowing a nerve impulse to trigger a molecular event within SM cells (myocytes) which gives rise to motion and force.

1.1.1 Skeletal muscle structure and function

SM is made up of bundled fascicles, which are themselves bundles of muscle fibres. Muscle fibres are large multi-nucleated cells rich in mitochondria and myofibrils. Myofibrils are units of the myocyte containing specialised proteins which make up the contractile units, sarcomeres, contained within myofibrils

(Figure 1.1).

Thick and *thin* myofilaments make up a myofibril. Thin filaments are comprised of the proteins actin and titin; thick filaments of the proteins myosin and nebulin. It is the 'climbing' of the head structure of myosin along the polymerised actin filaments, referred to as cross-bridge cycling, which gives rise to force generation. The mechanism by which this is thought to occur is referred to as the *sliding filament theory* (Figure 1.2). Crucially, this climbing action of the myosin head along the thin filament is controlled by the cytoplasmic availability of calcium ions because Ca^{2+} binding to troponin unblocks the actin binding site. Ca^{2+} availability is indirectly controlled by the conveyance of a nerve impulse in a process called excitation-contraction coupling (ECC).

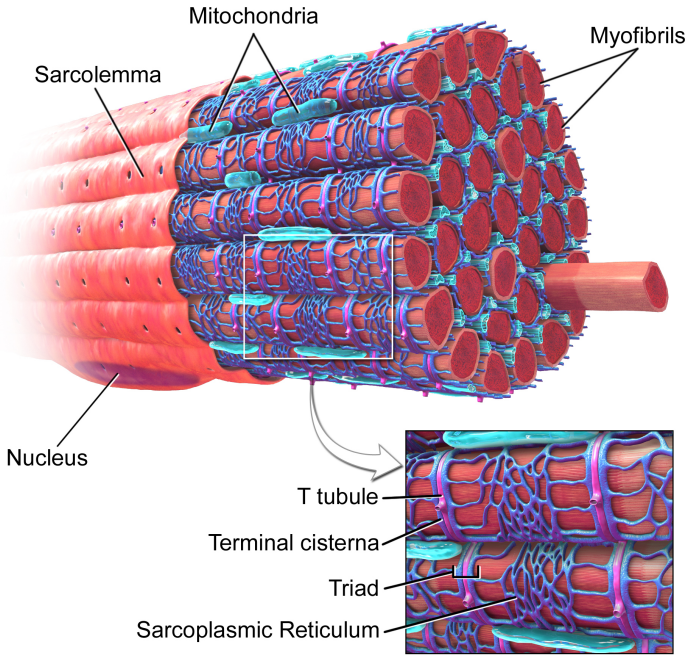

Figure 1.1: A schematic representation of a single muscle fibre
The muscle fibre contains bundles of myofibrils, protein-rich rod-like structures that contain the contractile units *sarcomeres*. Surrounding the myofibrils is a web-like organelle, the sarcoplasmic reticulum (SR). The SR holds and controls the release of Ca^{2+} into the cytoplasm. Ca^{2+} acts as a signalling molecule for a large number of cell processes, one of which is the initiation of cross-bridge cycling. Transverse tubules (T-tubule) are invaginations of the sarcolemma which assist in the propagation of an action potential to the SR. Figure from Blausen.com staff[1], distributed under a CC BY 3.0 license.

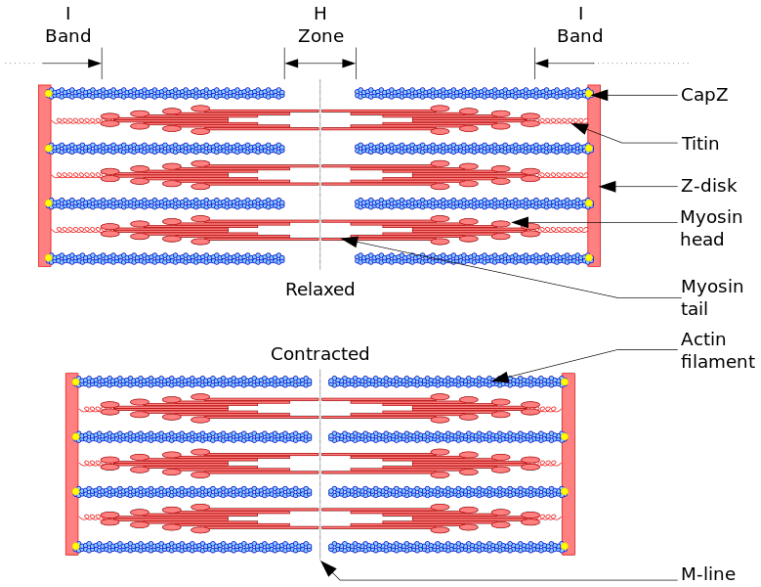

Figure 1.2: Schematic representation of a sarcomere and the sliding filament theory The sliding filament theory postulates that force generation is achieved through cross-bridge cycling within the contractile unit of the myocyte - the sarcomere. Cross-bridge cycling involves cyclic attachment and detachment of the myosin head to the actin filament, which pulls the actin filament toward the centre of the sarcomere. Ca^{2+} binding to troponin controls the attachment of the myosin head to actin, ultimately allowing the sarcomere to cycle between a relaxed and contracted state in a Ca^{2+}-dependent manner. Figure from David Richfield[2], distributed under a CC BY-SA 3.0 license.

1.1.2 Excitation-contraction coupling

A defining feature of myocytes is the presence of the sarcoplasmic reticulum (SR), a modified smooth endoplasmic reticulum. The SR contains intracellular stores of Ca^{2+}, the primary purpose of which is to allow for regulated and rapid increases in cytoplasmic Ca^{2+} to initiate cross-bridge cycling through troponin binding. The molecular events that result in Ca^{2+} release from the SR and subsequently muscle contraction are collectively termed excitation-contraction coupling (ECC).

At the neuromuscular junction, the contact point between a nerve and a myocyte, a nerve impulse is transmitted to the myocyte via a rapid release of acetylcholine (ACh) into the synaptic cleft of the nerve cell. ACh binds to ACh receptors of the plasma membrane (PM), termed the sarcolemma, of the myocyte. This triggers the opening of cation channels, leading to a rapid flow of

ions across the membrane, which causes a sudden depolarisation at the membrane surface. This depolarisation is said to generate an *action potential*, which travels down the transverse tubules of the sarcolemma. This is sensed by the dihydropyridine receptor (DHPR), an L-type membrane channel protein embedded in the T-tubule membrane, which undergoes a conformational change upon recognition of an action potential. Physical coupling between DHPR and the ryanodine receptor (RyR) Ca^{2+} channel leads to RyR channel opening upon the conformational change of DHPR[3,4]. RyR resides in the SR membrane, and its opening causes a release of Ca^{2+} from the SR into the cytoplasm to initiate muscle contraction. Aside from direct RyR accessory proteins to be discussed, the functional interaction between DHPR and RyR is modulated by Stac3 and JP-45 in skeletal muscle. These proteins are both thought to be auxiliary DHPR subunits, however their exact roles are yet to be fully understood. Stac3 seems to be required for ECC as determined by mouse experiments[5], while JP-45 is required for the correct functional expression of DHPR[6].

1.1.3 The ryanodine receptor

The ryanodine receptor is a very large, homotetrameric channel responsible for the passive conductance of Ca^{2+} from the SR. The muscle-specific RyR channel, RyR1, is encoded by the *RYR1* gene. The other isoforms of RyR, encoded by distinct genes, are RyR2, which is expressed in heart, and RyR3, expressed primarily in brain. The properties of each isoform vary and relate to their tissue-specific function[7], but the focus from here will be on RyR1.

A single RyR1 subunit is approximately 5000 amino acids in length, encoded by the 106-exon *RYR1* gene. The majority of the RyR1 channel resides cytoplasmically, with the remainder of the protein residing within the SR membrane and the SR lumen[8]. The size of RyR1 allows it to be regulated by a range of proteins and other effectors. Additionally, it is subject to a number of post-translational modifications (PTMs). This regulation results in tight control of RyR1-mediated Ca^{2+} release, allowing it to respond rapidly to changes in Ca^{2+} levels, as well as energy and reduction-oxidation (redox) states, of the myocyte (summarised in Figure 1.3).

1.1.3.1 RyR1 regulation

RyR1 is able to sense both cytoplasmic and SR Ca^{2+} levels through the Ca^{2+}-binding proteins calmodulin (CaM) and calsequestrin (CsQ), respectively. This subjects RyR1 to negative feedback regulation, so that increasing levels of cytoplasmic Ca^{2+}, and decreasing levels of SR Ca^{2+}, both lead to RyR1 channel closure. CaM is a cytoplasmic RyR1-interacting protein which inhibits RyR1 channel opening when Ca^{2+}-bound, but activates RyR1 when not Ca^{2+}-bound[9]. Additionally, an allosteric change of RyR1 is induced when it is bound to Ca^{2+} which may increase its affinity to CaM[10]. Another RyR1 cytoplasmic interacting Ca^{2+}-binding protein, S100A1, is thought to alter the CaM-bound structure and modulate RyR1 activity[11]. On the other hand, CsQ resides within the SR lumen and plays a comparable but opposite role to CaM by increasing the RyR1 open-channel state with increasing SR Ca^{2+} [12]. When CsQ's RyR1 regulatory function was first identified, it was discovered that this effect was dependent on other RyR1 binding partners[13]. These other important partners were determined to be triadin and junctin, SR-membrane associated proteins residing in the SR lumen. These proteins interact with RyR1 and CsQ, as well as each other[14]. Triadin is thought to anchor CsQ to the SR membrane, and junctin to modulate CsQ's inhibition of RyR1[15,16]. In mouse models, it has been found that through association of JP-45, CsQ may also indirectly modulate DHPR to allow negative regulation of ECC in response to depletion of SR Ca^{2+} stores[17].

Homer 1 and FKBP12 are two other proteins which regulate RyR1 activity in a non-Ca^{2+} dependent manner. Homer 1 is a direct interactor[18] and modulator of RyR1 channel activity[19]. Two different splice variants of the protein have been shown to activate or inhibit open channel probability in a dose-dependent manner[20]. FKBP12's role on the modulation of RyR1 is contentious, with research suggesting the protein stabilises both the open and closed state of RyR1 during ECC[21,22]. FKBP12 also acts in a pathway involving RyR1 regulation by phosphorylation in response to cAMP (cyclic adenosine monophosphate) levels: Increases in the concentration of cAMP leads to phosphorylation by protein kinase A of RyR1 which enhances the probability of RyR1 being in an open channel state[23]. Decreasing cAMP levels lead to RyR1 dephosphorylation by protein phosphatase 1 which allows binding of FKBP12, and this RyR1-FKBP12 interaction enhances the RyR1 closed channel state probability[23]. In agreement with what is known about the effect of RyR1 phosphorylation, hyperphosphorylation

of RyR1 has been associated with disease states in which there is a constitutive abnormal ('leaky') Ca^{2+} release from RyR1 in resting muscle[24,23]. FKBP12 binding to RyR1 is reduced by *S*-nitrosylation, another PTM of RyR1[25]. Via alterations in cellular redox states, *S*-nitrosylation of RyR1 occurs at physiological O_2 pressure (pO_2), which acts to activate RyR1. High pO_2 leads to removal of *S*-nitrosylation and activation of RyR1 and consequently an increase in muscle force production[26].

Other PTMs of RyR1 occur in response to changes in pO_2, allowing dynamic Ca^{2+} release dependent on the O_2 demand of muscle[26,27]. These are *S*-oxidation, *S*-palmitoylation and *S*-glutathionylation. All of the aforementioned PTMs occur via molecular pathways involving a number of proteins and molecular species: It is beyond the scope of this review to detail them. Additionally, the co-localisation of nitric oxide synthase, an enzyme which produces the signalling molecule nitric oxide, to RyR1 is associated with its nitration and increased Ca^{2+} release from the SR[28]. Whether or not RyR1 nitration is directly causative of its activation has not been established. Nevertheless, it suggests another mechanism by which RyR1 activity may be regulated.

Small allosteric effectors of RyR1 are Mg^{2+}, adenosine triphosphate (ATP), and as previously mentioned, Ca^{2+}. ATP and Ca^{2+} activate RyR1 by binding to activation sites[29,30]. Mg^{2+} binding exerts an inhibitory effect on RyR1 by both competing for binding at activation sites of ATP and Ca^{2+} and by binding to inhibitory sites[31,32].

1.1.4 Calcium homeostasis of myocytes

Ca^{2+} signalling by RyR1 to initiate muscle contraction has been discussed, but how is general Ca^{2+} homeostasis maintained in the myocyte? The flow of Ca^{2+} also occurs back into the SR from the cytoplasm, as well as between the cytoplasm and the extracellular spaces (summarised in Figure 1.4). Sarco/endoplasmic reticulum Ca^{2+} ATPase (SERCA) is the active transporter responsible for the reuptake of Ca^{2+} into the SR[34]. SERCA demands a phenomenal amount of the cell's energy: in mice it was determined that in resting muscle, to maintain low cytosolic Ca^{2+}, SERCA contributes to at least 40% of O_2 consumption[35]. Phospholambin and sarcolipin are understood to be direct regulators of SERCA[36,37]. Within the SR, CsQ acts as a Ca^{2+} buffer by binding to it, which reduces the 'workload' of SERCA by lowering the Ca^{2+} concentration of the SR.

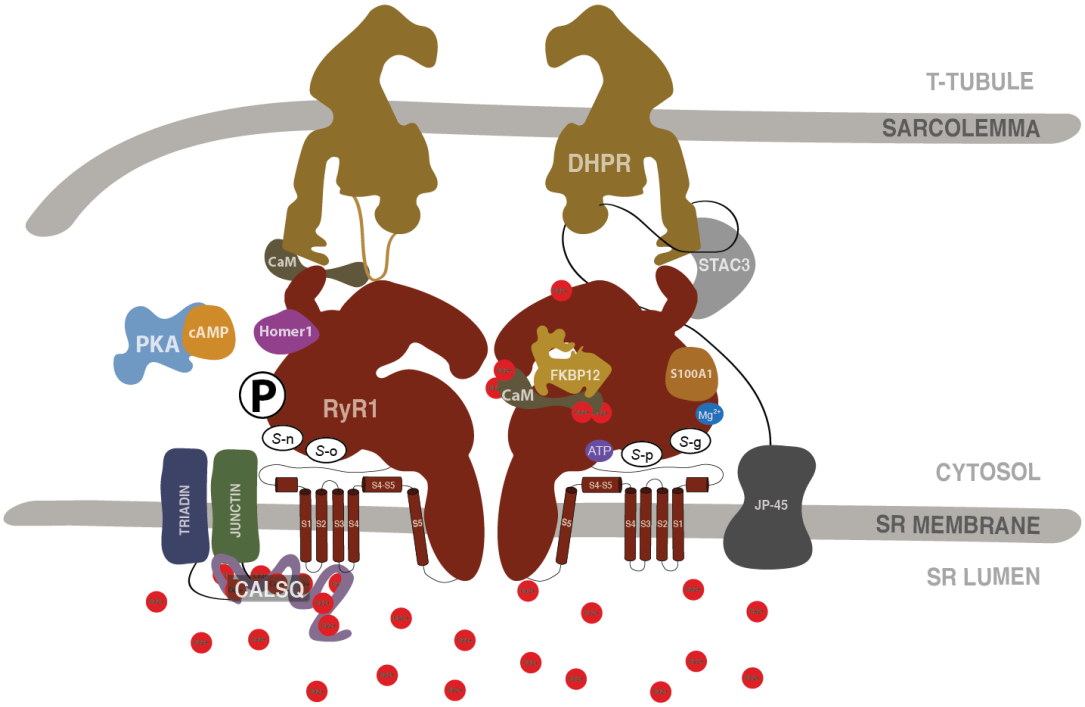

Figure 1.3: Schematic representation of RyR1 and its interacting proteins and ligands, as well as its post-translational modifications

The dihydropyridine receptor (DHPR), which transverses the sarcolemma, signals ryanodine receptor 1 (RyR1) opening by a conformational change upon recognition of an action potential[3,4]. Stac3 associates with DHPR and may be necessary for correct ECC[5]. JP-45 assists with the functional interaction between DHPR and RyR1[6], and may also associate with calsequestrin (CALSQ) to provide cross-talk between the ECC molecular machinery and sarcoplasmic reticulum (SR) Ca^{2+} levels[17]. Calmodulin (CaM) and CALSQ regulate RyR1 in response to Ca^{2+} concentrations of the cytoplasm and sarcoplasmic reticulum (SR), respectively[9,10,12]. S100A1 is a Ca^{2+}-binding protein which regulates RyR1 activity by modulating CaM's binding to RyR1[11]. Junctin and triadin are SR membrane-bound proteins which modulate the action of CALSQ[13,14,15,16]. Expression of two different Homer 1 isoforms differentially regulate RyR1 in a dose-dependent manner[18,19,20]. FKBP12 may stabilise both the open and closed RyR1 channel state[21,22]. FKBP12 binding occurs in RyR1's dephosphorylated state[23]: protein kinase A (PKA) phosphorylates RyR1 in response to increases in cyclic AMP (cAMP) concentration[23]. Redox state-dependent PTMs of RyR1 are *S*-nitrosylation (*S*-n), *S*-oxidation (*S*-o), *S*-palmitoylation (*S*-p) and *S*-glutathionylation (*S*-g) which allow RyR1 to be regulated by pO_2 of the myocyte[25,26,27]. ATP, Mg^{2+} and Ca^{2+} are allosteric effectors of RyR1[29,30,31,32]. Image is adapted from Sophie Burling with permission[33].

1.1.4.1 Store-operated calcium entry

Store-operated Ca^{2+} entry (SOCE) is responsible for the replenishment of SR Ca^{2+} stores through entry of extracellular Ca^{2+}. In SOCE, the protein stromal interaction molecule 1 (STIM1) senses and responds to decreasing Ca^{2+} levels of the SR lumen by translocating to the junction of the SR and PM, triggering the tetramerisation of Orai1, producing a functional Ca^{2+} channel which transverses both the PM and SR membrane and works to pump Ca^{2+} into the SR[38].

1.1.4.2 Calcium channels of the plasma membrane

A number of PM Ca^{2+} channels function to control Ca^{2+} flux between the cytoplasmic and extracellular spaces. NCX is a Na^{+}/Ca^{2+} antiporter which rectifies Ca^{2+} gradients during action potential generation through Ca^{2+} export[39]. The active transporter PMCA is also a Ca^{2+} export channel of the PM responsible for maintaining low cytoplasmic Ca^{2+} [39]. TRP (transient receptor potential) channels also contribute to Ca^{2+} homeostasis. This diverse family of cation channels mediates signalling events by flow of cations in response to a wide range of stimuli. TRPs are present within the PM and SR of skeletal muscle, with the subfamilies TRPC and TRPV being selective for Ca^{2+} [40]. Channels may be activated by temperature and certain ligands. Although the extent of the TRP channel's contribution to Ca^{2+} homeostasis within skeletal muscle is not currently fully understood, TRPC isoforms may significantly contribute to abnormal Ca^{2+} influx in the muscle disorder Duchenne muscular dystrophy[41].

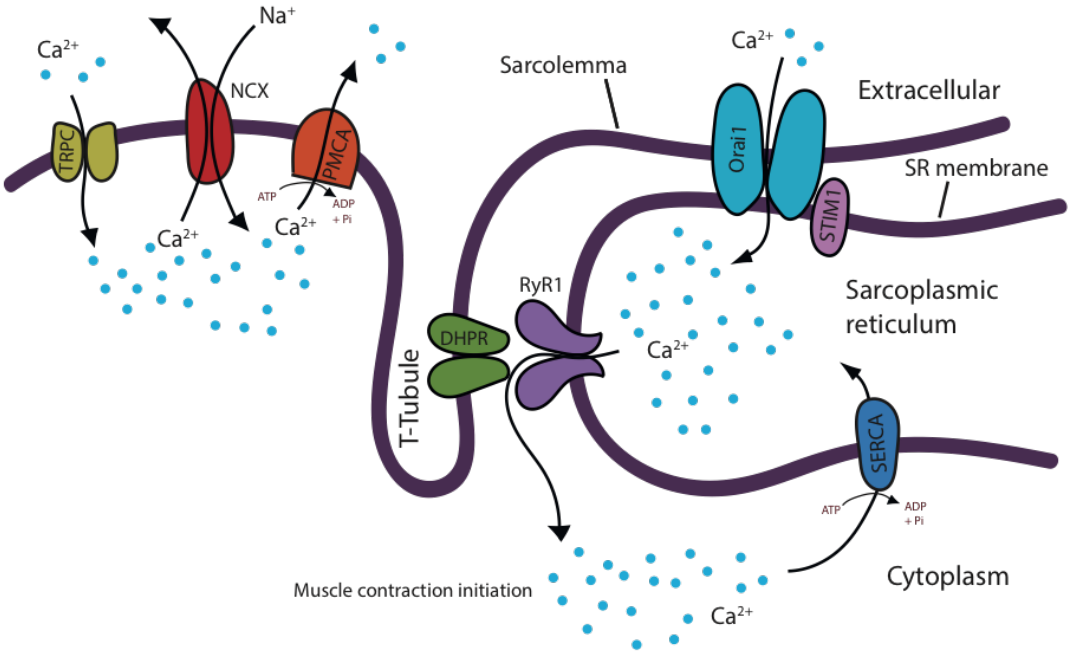

Figure 1.4: Schematic representation of ion channels contributing to Ca^{2+} homeostasis within the myocyte

Ca^{2+} is represented as blue circles; arrows represent direction of ion movement. Muscle contraction results from the generation of an action potential triggered by a nerve stimulation event. The action potential travels down transverse-tubules (T-tubule), invaginations of the sarcolemma. The dihydropyridine receptor (DHPR) responds to this by undergoing a conformational change. Physical coupling between DHPR and ryanodine receptor 1 (RyR1) leads to RyR1 channel opening which results in a passive flow of Ca^{2+} from the sarcoplasmic reticulum (SR) into the cytoplasm, where Ca^{2+} initiates muscle contraction through binding to troponin. Sarco/endoplasmic reticulum Ca^{2+} ATPase (SERCA) actively pumps Ca^{2+} back into the SR [34]. During prolonged muscle contraction SERCA Ca^{2+} transport alone cannot replenish SR Ca^{2+} stores and so the store-operated Ca^{2+} entry (SOCE) pathway, which involves the proteins stromal interaction molecule 1 (STIM1) and Orai1, is activated to allow extracellular Ca^{2+} uptake directly into the SR [38]. Ca^{2+} plasma membrane channels NCX and PMCA mediate Ca^{2+} transport over the plasma membrane (sarcolemma) in response to action potential generation and to ensure optimal Ca^{2+} levels are maintained within the cytoplasm [39]. Transient receptor potential channels (TRPC), of which there are a number of isoforms, are understood to uptake Ca^{2+} in response to a variety of signals and stimuli [40]. Original artwork.

1.2 Malignant hyperthermia: A pharmacogenetic disorder of muscle

Volatile anaesthetic drugs are given routinely for surgery and are generally considered safe. However, a person susceptible to the genetic disorder malignant hyperthermia (MH) is at risk of a hypermetabolic episode triggered by certain potent volatile anaesthetic agents[42]. An MH episode is life-threatening and must be treated immediately to prevent death. Secondary complications, such as rhabdomyolysis (breakdown of protein) leading to kidney failure, may occur. Drugs that may induce an MH episode in susceptible individuals are volatile anaesthetic compounds, including halothane and sevofluorane, as well as the depolarising muscle relaxant suxamethonium[42]. Non-volatile anaesthetic drugs should not trigger an MH episode in susceptible individuals. MH-susceptible (MHS) individuals are usually asymptomatic, although exercise- and heat-induced MH-like episodes have been reported[43]. An MHS individual may not react to general anaesthetics in the first instance: Several surgeries may be uneventful before an MH episode occurs[42].

An MH episode presents as muscle rigidity, increased CO_2 production, elevated blood K^+, hyperthermia and tachycardia[42]. On a cellular level, these symptoms are understood to be caused by dysregulation of Ca^{2+} homeostasis within myocytes[42]. This causes persistent muscle contraction, leading to sustained heat production and excess CO_2 production through oxidative metabolism. Depletion of ATP stores causes breakdown of myocytes, leaking K^+ and proteins into the bloodstream, with the latter causing rhabdomyolysis. Dantrolene is the only known effective treatment against a suspected MH episode as it is an antagonist of the Ca^{2+} release channel of myocytes, RyR1[42].

1.2.1 The genetics of malignant hyperthermia

MH is understood to be inherited in an autosomal dominant manner. Initial genetic work established that in some families MH-susceptibility was linked to chromosomal segment 19q12-13.1 (MHS1 locus), containing the gene *RYR1*[44]. Since then, according to European Malignant Hyperthermia Group (EMHG) guidelines, 48 variants [§] of *RYR1* have been accepted as pathogenic for MH[45],

[§]accepted EMHG variants available at https://www.emhg.org/diagnostic-mutations Hopkins et al.[45] contains EMHG guidelines for the acceptance of an MH-pathogenic variant.

with many more identified *RYR1* variants thought to be associated with MH-susceptibility. Aside from the MHS1 locus, other loci have been associated with MH-susceptibility: MHS2 (17q11.2-q24)[46], MHS3 (7q21-q22)[47], MHS4 (3q13)[48], MHS5 (1q32)[49] and MHS6 (5p)[49]. But only the MHS5 locus, which harbours the gene *CACNA1S* (encoding DHPR subunit $Ca_v1.1$), has been conclusively linked to MH-susceptibility. $Ca_v1.1$ directly interacts with RyR1[50]. Two *CACNA1S* variants have been classified as pathogenic for MH by the EMHG, with a number of other variants suspected to be pathogenic[51,52].

Over 50% of MH-susceptible families harbour a pathogenic *RYR1* or *CACNA1S* variant. A variant within the gene of the DHPR regulatory subunit, β_{1a}, has been linked to MH-susceptibility but is not currently accepted by the EMHG as pathogenic for MH[53]. MH has also been linked to another gene, *STAC3*, encoding the DHPR-associated protein Stac3: the variant is thought to cause Native American myopathy which predisposes sufferers to MH-susceptibility[54]. The genetic cause of MH in other families is currently unknown. If a genetic variant is suspected of being pathogenic for MH-susceptibility, segregation analysis is carried out to see if the variant follows the MH-susceptibility trait within a family. If segregation is established, a cohort of MH negative (MHN) individuals are then tested for the variant. If the variant is not present in any MHN individual, the variant is functionally characterised if possible. For *RYR1* variants, Ca^{2+}-release assays are carried out using mammalian cells transfected with *RYR1* cDNA containing the variant of interest[55,56]. Hypersensitivity of the RyR1 channel to RyR1 agonists can confirm that the suspected variant is likely to lead to an MH-susceptible phenotype. At present, these Ca^{2+} release assays have not been successfully carried out on other genes of interest to MH research due to the difficulty in culturing myocytes which act functionally the same as *in vivo*.

Prevalence estimates of MH vary, and depend on the manner in which the disorder's presence is defined. MH-pathogenic variant prevalence is estimated at around one in 800 individuals[57], whereas the actual number of suspected MH episodes may be between one in 10,000 to 250,000 anaesthetics[58,59]. Of course, due to the genetic nature of the disorder, there are MH 'hot spot' regions. For example, in Palmerston North, New Zealand, approximately one in 175 persons receiving general anaesthetics are treated as MH-susceptible due to the presence of a few very large founder families within the region.

1.2.2 Testing for malignant hyperthermia

The current accepted, gold-standard diagnostic method for MH is the *in vitro* contracture test (IVCT). The IVCT involves exposing a sample of skeletal muscle tissue taken from the quadriceps muscle of a patient, approximately 4 cm long and 1 cm deep, to increasing concentrations of contracture-triggering agents halothane and caffeine[60]. The contractile response of the muscle tissue to these triggering agents is subsequently measured; a positive IVCT result is recorded when a specific contractile response occurs above a threshold concentration of halothane or caffeine. Contractile response is measured in grams of force. A person with a positive IVCT result for both caffeine and halothane is labelled MHS_{hc} (from here also referred to as MHS), whereas a positive test for only halothane or caffeine is termed MHS_h or MHS_c, respectively. However, the latter two are treated as an MH-susceptible diagnosis for clinical purposes and as such should not be given MH-triggering anaesthetics. A person with a negative IVCT test result is labelled MHN and can safely receive volatile anaesthetic agents.

1.2.2.1 Threshold contracture values

The threshold values for caffeine and halothane can differ among IVCT testing facilities because each facility has their own 'in-house' controls[61]. For negative (MHN) controls, facilities perform the IVCT on leftover muscle from orthopedic surgeries on individuals with normal, non-MH-like anaesthetic responses. Therefore the corresponding contracture values for individuals diagnosed as MHS, MHS_h, MHS_c and MHN may differ. With this being said, in Australia, contractile responses of over 0.2 g for halothane and caffeine are diagnosed MHS[45]. In New Zealand, this threshold is 0.4 g for halothane and 0.2 g for caffeine[61]. IVCT response values at or near these thresholds are regularly referred to as 'borderline', and due to the potential for variability in IVCT result diagnoses between testing facilities, these values could be described as falling within an IVCT contractile response 'grey area'. This has important implications for the identification of MH-pathogenic variants within a family due to the possibility that an individual diagnosed as MHS with borderline IVCT values may not actually harbour a variant pathogenic for MH.

1.2.2.2 Limitations of the IVCT

The IVCT is both costly to the healthcare system and invasive to the patient. Perhaps its largest flaw is that it has a relatively low specificity rate (the rate at which MH negative individuals are labelled MH-susceptible), estimated at 6%[62]. This low specificity ensures that MHS individuals are not misdiagnosed, as with it comes a very high sensitivity rate (the rate at which MHS individuals are labelled MHN) of <1%[62]. One major goal of MH research is to find a more effective diagnostic test; determining the genetic cause of MH may allow the IVCT to be superseded by DNA-based diagnostic methods. Unfortunately, misdiagnosis of MH-susceptibility due to the inaccuracy of the IVCT is likely to be hindering efforts to identify the genetic cause of MH.

1.2.3 The current biochemical understanding of malignant hyperthermia

It is well established that MH is caused by abnormal and excessive Ca^{2+} release from the SR. There is a loss of normal Ca^{2+}-release negative feedback regulation. This is caused specifically by volatile anaesthetic agents, and additionally by the PM depolarising muscle relaxant suxamethonium. It is obvious that an understanding of the mechanism of action of these pharmaceutical agents would help to understand the pathomechanisms of MH. Unfortunately, the mechanisms of action of volatile anaesthetics are currently mostly unknown. More is known about the effect of suxamethonium however: It competes with ACh for binding to the ACh receptor at the synaptic cleft of the myocyte. Normally, acetylcholine is released from the receptor upon binding and quickly hydrolysed, allowing another successive round of acetylcholine release from the motor neuron and binding to the receptor. However, suxamethonium is not hydrolysed and has an extended effect, disallowing repolarisation of the myocyte to occur. How exactly suxamethonium action causes RyR1 channel opening is unknown. What is known is that *RYR1* variants that have been functionally characterised *in vitro* as well as *ex vivo* using patient tissue samples cause a destabilisation of the RyR1 closed channel state, making RyR1 hypersensitive to activation upon exposure to triggering agents[56, 55, 42].

Based on what is known about MH aetiology, it is plausible that unknown MH-pathogenic variants will be present within genes encoding proteins with a role

in ECC. The variant could lead to a dysfunction of the protein, causing RyR1 hypersensitivity. Although suspicion has long been cast on genes with a role in ECC, currently no variants of these genes other than those falling within *RYR1*, *CACNA1S* and *STAC3* have been linked to MH. Interestingly, the role of the protein Stac3 was only recently identified, suggesting that there could certainly be other currently unknown proteins important for ECC and/or RyR1 regulation. These could themselves be core components of the ECC multi-protein complex or direct RyR1 modulators, but alternatively could be important for the functional expression of ECC components or RyR1 modulators.

There are a number of other possibilities. It could be that an unknown MH-pathogenic variant falls within a gene encoding a protein involved in the PTM of RyR1, either directly or indirectly. Acute or chronic alterations in Ca^{2+}, redox or energy states may lead to PTMs of RyR1. Incorrect PTM of RyR1 could cause the channel to become hypersensitive. Alternatively, the pathogenesis may involve another myocyte Ca^{2+} channel or a protein involved in the regulation of a Ca^{2+} channel. As there is much to be understood about MH, and there are many families with yet-to-be discovered MH-pathogenic variants, these possibilities, and others, will have to be explored if the cause of MH is to be fully understood.

1.2.4 Malignant hyperthermia and other diseases

There are well-established links between MH-susceptibility and a number of muscle diseases (myopathies). The most closely coupled to MH-susceptibility is central core disease (CCD), a non-degenerative myopathy causing muscle weakness and MH-susceptibility[63]. It presents histologically as muscle fibres containing 'cores': Regions with reduced mitochondrial activity spanning longitudinally down the muscle fibre[63]. *RYR1* variants are most commonly the cause of CCD[64]. However, these variants have been characterised as leading to 'leaky', rather than hypersensitive, RyR1 channels, whereas MH-pathogenic *RYR1* variants are characterised as hypersensitive and sometimes leaky[65,66]. 'Leaky' variants lead to a constant leak of Ca^{2+} which is thought to cause muscle weakness due to a decrease of Ca^{2+} in the SR.

Other myopathies which may predispose individuals to MH-susceptibility are King-Denborough syndrome, multi-minicore disease and central nuclear myopathy, with *RYR1* variants also being linked to these disorders[67,68,69,70].

Tubular aggregate myopathy (TAM) is a disease which is characterised by

atrophy of type II muscle fibres and formation of tubular aggregates histologically visible under the microscope. TAM sufferers may be at risk for MH[71]. TAM has been linked to STIM1 gain-of-function mutations which lead to constitutive flow of extracellular Ca^{2+} into the SR[72], but it is currently unclear how this leads to the TAM phenotype.

MDMA-induced hyperthermia is an abnormal, often lethal metabolic episode triggered by 3,4-Methylenedioxymethamphetamine (MDMA) which has symptoms mirroring those of an MH episode: muscle contraction, metabolic acidosis and rhabdomyolysis[73]. Importantly however, the mechanism by which MDMA-induced hyperthermia is triggered is thought to differ from MH: It has been hypothesised that it is mediated by a hypothalamic response which releases norepinephrine and by an increase in thyroid hormone, both triggered by MDMA, which leads to hyper-activation of the myocyte mitochondrial uncoupling-protein 3 (UCP3)[74,75,76]. UCP3 mediates passive proton leakage across the inner mitochondrial membrane causing a futile cycle which expends ATP and produces heat, causing the symptoms of an MDMA-induced hyperthermia episode[74]. Despite this, the similarity between the conditions is certainly noteable.

1.3 Whole exome sequencing for the discovery of disease-causative genetic variants

Improvements in the quality and affordability of next-generation sequencing (NGS) methods have rapidly changed the way in which genetic diseases are researched. When trying to identify the cause of a heterogeneous genetic disease or a genetic disease with unknown aetiology, NGS is now often carried out in the first instance. This has sped up the rate of discovery of previously-elusive causes of genetic diseases[77].

Currently, 1-2% of the human genome is thought to contain protein-coding genes[78]; the remaining genome, aside from non protein-coding genes which play important roles in the control of gene expression, is sometimes labelled as 'junk' DNA, although there is growing evidence that this DNA plays a vital role in chromatin structure and gene regulation[79]. It is believed that pathogenic variation in the protein-coding regions (the exome) of the genome contributes to the vast majority of genetic disorders.

1.3.1 The principle of whole exome sequencing

Whole exome sequencing (WES) is a targeted alternative to whole genome sequencing (WGS) which involves capture and enrichment of the exonic genomic regions prior to sequencing. To achieve this, biotinylated RNA or DNA probes (also called *baits*) with sequences corresponding to exonic DNA sequences are used to pull-down (capture) the exonic DNA from adaptor-ligated, fragmented gDNA. This is followed by amplification using the polymerase chain reaction (PCR) which results in a pool of DNA which should represent the known exome. NGS methods are then applied to sequence the amplified exome. Other targeted approaches to sequencing include *gene panels*, where probes corresponding to specific target genes are used to sequence a more limited set of genes of interest.

Why is WES chosen over WGS? Although WGS allows examination of the entire genome, there are two main reasons that WES is carried out as an alternative to WGS for genetic disease studies: The first reason is that it is less expensive, being approximately four-fold cheaper than WGS[80]; the second reason is that only the region of the genome most likely to harbour pathogenic variation is sequenced, creating much less data, meaning less computation time and analysis (which both come at a cost). Interpretation of variation within the exome

is also considerably more manageable. Despite these reasons, studies have found that WGS is better able to detect variation of the exome compared to WES[81,82]. However, these studies are not entirely relevant today due to improvements in WES technology in recent years, with more recent studies suggesting that disparities between WGS and WES in terms of exome variant detection are minimal and may constitute $\approx 1\%$ of the human exome[83,84].

1.3.2 Weaknesses of WES

Although WES is often chosen in genetic disease research for the reasons described above, it does have drawbacks. Any sequencing step which involves capture and/or amplification of nucleic acid leads to the potential for sequencing bias, and this is certainly the case for WES. Allele distribution bias and allelic dropout are seen in WES data due to both capture and PCR amplification steps, which may prevent variant identification during computational processing. For the capture step, both polymorphic and rare variation of the genome may inhibit or completely prevent probe binding[80], leading to hidden heterozygosity or no sequence read data in the region containing variation. Due to PCR's inherent stochastic processes, the same loss of information may also be seen[85]. Additionally, there is likely to be a prejudice in WES kit probe design against less-represented genetic populations, especially indigenous populations. It has been suggested that WES kit probes be designed to alternative haplotypes to reduce this bias[80]. Extremes of GC (guanine/cytosine) sequences are problematic for WES due to the necessity of PCR amplification[80,86,82,81]. PCR also reduces the ability to identify copy-number variation (CNV), however, a number of programmes designed specifically to detect CNVs in targeted sequencing data have recently emerged and are leading to improvements in the diagnostic yield of WES[87,88].

Overall, it is evident that if cost and computational resources are not factors, WGS is a better option. But currently, WES is the preferred option for genetic disease studies. As WGS becomes more affordable and the necessary computational infrastructure more accessible, WGS will undoubtedly become the norm. Indeed, although there is currently incentive for genetic disease researchers to choose WES over WGS, this could be preventing important advancements in the understanding of genetic disease caused by non-coding variation or CNVs.

1.3.3 WES kits

There are a number of commercially available WES kits (specific technologies and probe designs used to carry out WES sequencing). The major WES kit manufacturers are Nimblegen (now Roche), Illumina, and Agilent, producing SeqCap EZ Exome Library (Nimblegen), TruSeq Exome Enrichment Kit (Illumina), Nextera Rapid Capture Exome Kit (Illumina) and Sure Select Human All Exon Kit (Agilent), respectively. These WES kits differ in their probe type, strategy of target capture, and genomic target regions[89], making some better suited for certain research applications. For example, Agilent's Sure Select Human All Exon Kit uses RNA probes approximately 120 nucleotides in length which bind to the target adjacent to one another, whereas Nimblegen's SeqCap EZ Exome Library kit uses higher-density, variable-length DNA probes which overlap against the target region[86]. As summarised in Warr *et al.* (2015)[89], these distinctive approaches of exome capture culminate in different strengths and weaknesses of each kit. For example, Agilent's design is superior for small indel (insertion/deletion) detection, whereas Nimblegen tends to capture more SNVs (single nucleotide variants)[90]. Additionally, the WES kits are designed using different transcriptome databases: as discussed in Section 1.4.4, there is no clear consensus on what constitutes the human exome. This means that the genomic target regions of WES kits differ, resulting in variants that will be detected by some kits but not others[90]. It is important to note that the research presented by Chilamakuri *et al.*[90] in 2014 is likely outdated; more recent studies in this area were not identified. WES kit designs have undoubtedly changed and improved, and indeed the data presented in this thesis suggests that (see Chapter 3.2.2). However, it is important to acknowledge the differences in WES kits, especially if WES data from multiple WES kits will be pooled together in a genetic disease study.

1.4 Bioinformatic tools for variant discovery

With an increase in the ability to perform NGS has come a substantial increase in the volume of NGS data, and with this has come a vast number of tools to analyse these data. Although there are commercial options available, many tools, or programmes, are freely available for each step of NGS data processing. Here, the general computational steps involved in a bioinformatic pipeline for variant discovery in genetic disease research will be explained, and a selection of these

freely available tools will be presented and discussed.

1.4.1 Pre-processing of reads

Raw sequence read data is trimmed to remove adaptor sequences and, if applicable, de-multiplexed to remove sample-specific sequence tags. This is followed by performing quality-control on the reads, which involves removing reads that do not meet pre-defined criteria in terms of read quality. This results in two (if paired-end sequencing has been carried out as is often the case) analysis-ready *fastq* files per sample. Fastq files also contain encoded information, obtained from the sequencing machine, about the quality of each called nucleotide.

1.4.2 Sequence alignment

Sequence alignment (also called *mapping*) is carried out to align the reads from the fastq file to a reference genome sequence. This results in a *bam* file, which contains binary information about which position of the reference genome (if any) each sequence read from the fastq file has aligned to. Bam files also encode mapping quality scores, which give an indication of the likelihood that each read has been mapped to the genomic position from which it originated. Coverage statistics may be calculated using bam files. *Coverage* is defined as the number of reads aligned to a given base or genomic region. The term coverage is synonymous with *read depth*, or *depth*. Coverage can provide an indication of the success of sequence alignment, and by extension, the quality of NGS data[91].

There are two main sequence alignment programmes commonly used for human NGS data: BWA-MEM[92] (burrows-wheeler aligner) and bowtie2[93]. These programmes differ in the algorithms that they employ to carry out sequence alignment. BWA-MEM is designed for short-read alignment of sequences of high similarity, such as in the case of human sequence reads against the human reference genome, and is a popular choice of aligner for WES[92].

1.4.2.1 Human reference genome

Since the human genome was first sequenced, there have been multiple releases of the human reference genome. Hg38, released in 2013, is the current human reference genome major release. It is beneficial to use the most recent release as it constitutes the most current understanding of the sequence of the

human genome. Hg38 includes alternative loci in genomic regions of high variability between individuals and populations. Despite this, the previously released reference genome, hg37, is still in use today, as many tools and programmes are only compatible with hg37.

1.4.3 Variant calling

Variant calling involves making inferences about the presence of a variant based on mismatches between sequence reads and the reference genome. Variant calling programmes 'call' a variant by taking into account the quality of the sequence read data, the mapping quality scores, and the proportion of reads containing the mismatch. A list of called variants is given, along with a 'confidence' score (quality score) representing the probability that the variant is actually present, in the form of a *VCF* (variant call format). Variants can be filtered based on this quality score. This is termed *hard filtering*, as it involves applying a threshold to remove variants that do no meet a certain quality score. Applying hard-filtering to remove false-positive variants will come at the expense of losing true variants, and so the thresholds applied in hard filtering are dependent on the purpose of the study. For example, it may be beneficial to have a low stringency when searching for disease-causative genetic variants to ensure variants of interest are not missed.

Current popular variant callers include the Genome Analysis Toolkit (GATK) HaplotyperCaller[94], Strelka2[95], Samtools[96] and FreeBayes[97]. Because these programmes differ in their method of variant calling, there is some lack of concordance between them, with some outperforming others in certain areas[98,99,100]. For example, GATK HaplotyperCaller has been found to be superior for indel identification, but has under-performed in SNV identification when compared to Samtools[100,99] and Freebayes[99]. Despite this, GATK HaplotypeCaller appears to be the most popular variant calling platform for WES studies currently. Strelka2 is a relatively recent programme, released in 2018, which is suggested by the developers to outperform GATK HaplotypeCaller in terms of both speed and variant calling accuracy[95]. A recent comparison of these two variant calling platforms drew the same conclusions[94]. Overall, it is important to understand that variant calling platforms may yield different results with the same data, and as suggested by O'Rawe *et al.* (2013)[98], it may be beneficial to use multiple variant calling pipelines to reduce the chance of missing a disease-causative variant.

1.4.4 Variant annotation

Following variant calling, variant annotation is carried out to predict the effect that a variant may have, in turn allowing variants to be prioritised. Broadly, there are two types of variant annotation: the first and more fundamental type is *categorisation*; and the second is *functional effect prediction*. Both of these will be discussed.

1.4.4.1 Variant categorisation

Variant categorisation involves identifying whether a variant falls within a known or predicted gene, and whether it is predicted to change the gene's encoded product or expression of the gene (e.g. exonic, intronic, missense, nonsense, synonymous, splice-site acceptor, etc.). Categorisation of a variant is of course reliant on what is known of the set of human genes and the human transcriptome. There are two widely used datasets which contain a collection of known or predicted genes and transcripts: Ensembl[101] and RefSeq[102]. Although these datasets contain many of the same genes and transcripts, they differ to a significant extent. These differences can be attributed to the level of stringency applied by Ensembl and Refseq to accept a predicted gene or transcript into the dataset. Although Refseq is a smaller dataset than Ensembl, it is not simply a subset of the Ensembl dataset[103]. An important study highlighted the somewhat extreme effect these differences can have on variant categorisation. The research, which analysed the outcome of variant categorisation using both Ensembl and Refseq datasets, found only 44% concordance in predicted loss of function (LOF) variants[103], which are of great interest in genetic disease research.

Popular variant annotation software includes ANNOVAR[104], Variant Effect Predictor (VEP)[105] and SnpEff[106]. These programmes apply both categorisation and functional effect prediction annotations to variants. As well as the differences found between transcript datasets, McCarthy *et al.*[103] found significant differences in annotations applied to variants when comparing ANNOVAR and VEP.

1.4.4.2 Variant functional annotation

There are multiple types of variant functional annotations. A common type of annotation is to determine a variant's frequency within the general population, or subsets of the population. This is termed the allele frequency (AF). There are a

number of databases containing human AFs including 1000 Genomes[107] and gnomAD[108] (genome aggregation database). Currently, the largest of the common databases is gnomAD, with 140,000 human genomes and exomes. Using the AF in genetic disease studies can be a powerful method to filter variants: when the gnomAD database was initially released, it was determined that 90% of variants previously reported to be likely-pathogenic by genetic disease researchers were present at frequencies too high to be causative of genetic disease[109]. However, although the AF can help to rule out variants unlikely to be causative of a rare genetic disease, it is not possible to determine a variant's likely pathogenicity based on an AF alone. This is because the majority of variants identified in human NGS experiments will be rare (defined as $< 0.5\%$)[110]. In most AF databases including gnomAD, variant data is obtained from individuals who do not suffer from disease. However, it is possible that data from individuals with an undiagnosed genetic disorder, such as MH-susceptibility, may be present. Often these databases, as in the case of gnomAD, have AFs for ethnicity-specific subsets of variant data. Currently, European individuals are over-represented in the data. Noteably, AFs are not presented for Māori and Pacific Islander populations, likely owing to a lack of genomic data collected from these populations. This has implications for genetic disease research in New Zealand as variants which are common and non-pathogenic within individuals from Māori and Pacific Islander populations may appear novel or very rare in AF databases, which may mislead research directions. There is potential for greater representation of other populations; recently, the GenomeAsia 100K Project was launched, with plans to sequence a large number of genomes from Asian populations[111].

A more intricate but highly useful functional annotation is pathogenicity prediction (prediction of deleteriousness), for which there are many types. Predictions may be based on the conservation of a nucleotide or amino acid sequence at the variant position between species or amongst human genomes, the amino acid change, or *in silico* protein structural predictions. More recently, *ensemble* methods of pathogenicity prediction, which gather predictions from a range of tools, combining the power of the aforementioned pathogenicity prediction types as well as AFs, have become more popular. Commonly used ensemble tools include CADD[112] and REVEL[113]. CADD combines 60 different pathogenicity tool annotations into a single 'CADD score'. CADD scores are derived in the following manner: Using pathogenicity annotations, all possible SNVs within the

human genome are ranked against one another from least likely to be pathogenic to most likely. Raw rankings for all SNVs are then scaled and normalised to all other SNV scores to obtain scaled scores which represent the estimated deleteriousness relative to all other potential SNVs. For example, a score of 10 means the SNV is in the top 10% of SNVs, a score of 20 within the top 1% of SNVs, a score of 30 in the top 0.1% of SNVs, and so on. Often, a CADD score of over 15 is classed as being potentially pathogenic[114], but there is no strict definition. REVEL combines 13 different pathogenicity tool annotations[113]; it differs from CADD in that the developers focused on combining pathogenicity prediction annotations adept at sorting rare neutral variants from rare deleterious mutations. Their own benchmarking suggested REVEL to be superior to CADD and other prediction methods for identifying pathogenic variation[113]. Another recent study confirmed this finding but also determined that CADD tends to have a higher sensitivity (ability to identify truly pathogenic variants) compared to REVEL and other pathogenicity prediction tools[115]. One study which compared pathogenicity predictions of known MH-pathogenic variants found that CADD identified all known MH-pathogenic variants as likely pathogenic (using a CADD score cut-off of 15) although it had very low specificity compared to other tools tested[114]. It is important to note that pathogenicity predictions may be useful for the prioritisation of variants but cannot be relied upon to determine whether a variant is pathogenic or not.

Annotation of gene function is also useful. Gene Ontology (GO)[116,117] terms, which give information about gene function, can be applied to variants. Similarly, Online Mendelian Inheritance in Man (OMIM)[118] is an online database containing useful information about human genetic disease and genes linked to genetic disease. OMIM accessions can be applied to variants to identify whether the gene has been previously associated with genetic disease.

dbSNP[119], a public archive of all known human genetic variation, can be used to determine what is known about an identified variant. Each variant present in this database has a dbSNP identifier which can be used to quickly determine the AF of a variant and whether the variant has previously appeared in the scientific literature.

To determine whether a gene is expressed in tissues of interest, gene expression databases are available. These include GTEX[120] and the Human Protein Atlas[121]. The Human Protein Atlas provides GTEX expression data as well as

its own expression data. Gene expression data must be used with some caution however. For example, a search for the gene encoding FKBP12 (*FKBP1A*) on the Human Protein Atlas suggests that there is very low, or no expression, of this gene within skeletal muscle. Yet this protein has a well-defined role in RyR1 modulation within skeletal muscle[21,22]. Additionally, genes expressed early in muscle differentiation won't necessarily be expressed in a mature myocyte. Yet this does not preclude the gene from being of interest to a muscle disorder, as it is possible that a variant in a gene involved in muscle maturation could be pathogenic to the mature myocyte. As an example, the protein Dusp4 has recently been found to be implicated in correct muscle differentiation[122], yet is not expressed in mature skeletal muscle. If gene expression data were relied upon absolutely when carrying out muscle disease research, a variant within the gene encoding Dusp4 may be disregarded.

1.4.4.3 Variant filtering and prioritisation

Variants may be filtered based on variant annotation and other information, such as the phenotype of individuals harbouring the variant. There are a number of programmes designed for filtering, such as Variant Annotation and Filter Tool (VarAFT)[123] and GEnome MINIng (GEMINI)[124]. VarAFT is a graphical user interface programme whilst GEMINI is a command line programme, making VarAFT more user-friendly for researchers not experienced with the command line. Both of these programmes allow variant filtering based on an inheritance model (e.g. autosomal dominant, X-linked recessive) as well as filtering based on variant category, AF and pathogenicity prediction. VarAFT integrates gene expression data from the GTEX database and KEGG[125] to allow filtering based on gene expressivity and function.

1.4.5 Detection of copy-number variation

In comparison to WGS, the ability to detect copy-number variation (CNV) using WES data is reduced due to the necessity of target capture. Target capture increases the variability of coverage in WES, making it harder to determine whether an exonic region with significant deviation from the mean coverage is actually the result of a CNV. Nevertheless, there are a number of programmes designed to detect CNVs using WES data. These include CNVkit[126] and CODEX2[127].

1.4.6 Tool installation and use

The installation and use of many bioinformatic tools can be challenging, owing to the fact that many of these require use of the command line. This is a major hurdle for researchers, especially those who are not experienced with bioinformatic analyses. Even for experienced bioinformaticians, installation of tools can be difficult: One study found that up to 50% of randomly selected bioinformatics tools were either difficult to install (could not be installed within 15 minutes) or impossible to install (defined as not being installed within 2 hours)[128]. This can mean that even if suitable bioinformatics tools are selected for a project, installation may not be possible.

1.4.7 Which tools are best to use?

This review should make apparent the wealth of tools and resources available for WES data analysis for the identification of variants of interest in genetic disease studies. There are also a large number of other popular tools which have not been addressed here. Combinations of these tools result in an enormous number of potential bioinformatic pipelines. Although research has been undertaken to address which pipelines are superior for certain tasks, due to the ever-evolving nature of bioinformatic tools and resources, this research is likely to be quickly invalidated. If time and cost were no obstacle, the most effective option would be to run WES data through a number of pipelines to reduce the risk of missing a genetic variant causative of the disease in question. But as time and cost are important considerations, small differences in the performance of WES kits and bioinformatic pipelines may have to be disregarded, especially if difficulties in tool installation and use are encountered for the chosen pipeline. However, there are a number of easily implemented actions that may decrease the potential for missing variants of interest. For example, the variant annotation and filtering programme VarAFT, which uses ANNOVAR, applies both Ensembl and Refseq annotations, which are known to differ significantly, to variants.

What is likely to be useful is the re-analysis of WES data as the disease in question becomes better understood, and as more powerful bioinformatic tools and resources become available. Confirming this, a recent study concerning rare human genetic disease found that yearly re-analysis using the most recent bioinformatic resources and tools is diagnostically useful[129].

1.5 Project background

This project set out to determine the underlying genetic basis of MH in six families. In these families, variants with potential to cause MH within genes *RYR1*, *CACNA1S* and *STAC3* had previously not been identified, or identified variants had been ruled out as being the sole causative agent of MH. Previously obtained WES data from each family was used in an attempt to identify variants with potential to be pathogenic for MH.

1.5.1 Significance of study

Currently, the IVCT is the only available pre-symptomatic MH diagnostic method for families where familial pathogenic variants have not been identified. This method involves an expensive, time-consuming and potentially traumatic muscle biopsy. Aside from the invasiveness of the procedure, the IVCT is not considered completely specific, with a significant false-positive rate[62]. This means that individuals may be falsely diagnosed with MH-susceptibility, potentially leading to unnecessary future medical intervention. The overarching aim of MH genetic research is thus to achieve a DNA-based diagnostic test, and this firstly requires the identification of the pathogenic genetic variant in an MH-susceptible family. Understanding of the genetics and biochemistry of the disorder may also further understanding of myocyte physiology and biochemistry, as well as the aetiology of muscle disease. Additionally, this project aimed to establish bioinformatic methods with the potential to identify genetic variants of interest within families with MH. These methods could also be used for other complex genetic disorders with currently unknown genetic origin.

1.5.2 Hypothesis of study

Rare genetic variation present within the exonic region of the genome, leading to a protein sequence or protein expression change, is linked to MH-susceptibility in each family under study.

1.5.3 Aims of study

- To develop a bioinformatic pipeline that: efficiently processes whole exome sequencing data to obtain genetic variants present within each family under study; and enables effective and dynamic filtering and prioritisation of variants of interest.
- To select candidate variants that may contribute to the presence of MH-susceptibility in each family.
- To carry out genotyping within each family to establish whether there is a correlation between a variant of interest and MH-susceptibility.

1.5.4 Assumptions of study

1.5.4.1 Cause of MH

In this study, it is assumed that the pathogenic genetic variant in each family is present in the exonic region of the genome. Additionally, unless stated otherwise, it is assumed that MH-susceptibility in each family is due to a single genetic variant.

1.5.4.2 Accuracy of phenotypic data

One assumption is that the phenotypic data is accurate. To obtain phenotypic data for each individual, data transfer steps included manual recording to an excel spreadsheet from a hospital database followed by handwritten recording by the supervisor of this project. Handwritten records were then obtained and recorded in a separate spreadsheet by the student conducting the research. This was necessary for two reasons: Firstly, access to the MidCentral DHB database is strongly restricted and does not allow mass information download; secondly, these steps ensured decoupling of DNA numbers from individual identifying information. Although caution and accuracy was exercised during data transfer and recording, each of these steps did increase the potential for human error.

Additionally, as discussed in Chapter 1.2.2, results of the IVCT cannot be completely relied upon to distinguish actual MH-susceptible individuals from MH negative individuals.

1.6 Families

The six families under study are labelled A, B, C, D, E and F. In all of these families MH is inherited in an autosomal dominant manner. WES data from MHS_{hc}, MHS_h, and MHN individuals belonging to these six families were obtained to try to identify a potentially pathogenic genetic variant. gDNA and IVCT results were available for other individuals from these families, with the exception of family F, allowing for segregation analysis on identified variants of interest. Due to phenotypic complexities, variant analysis was carried out in a unique and discretionary manner for each family. For example, some MHS and MHN individuals obtained borderline IVCT values meaning it was considered possible they did not, or did, carry a pathogenic MH variant regardless of their diagnosis. Additionally, some families suffer from a myopathy additional to MH-susceptibility. For these reasons and others, the strategy for variant analysis was modified for each family accordingly (see Table 1.1). Each family and their specific considerations are described in detail below. A table containing all individuals for which WES samples were obtained, and their IVCT results, are available in Appendix A.33.

Consideration	**Adaptation to general strategy**
MHS IVCT result borderline or MHS_h/MHS_c	These individuals may not carry a potentially pathogenic variant
Presence of a known or suspected MH-pathogenic variant within *RYR1* or *CACNA1S*	MHS individuals carrying these variants may not carry the other potentially pathogenic variant
Presence of a myopathy	Increases unknown variant's potential to be present in a gene which may not normally be considered to have involvement in MH-aetiology
Large and branched family	Increases the possibility of the presence of a second potentially pathogenic variant

Table 1.1: Examples of the considerations taken with all families and consequently the adaptations made to the general strategy of variant analysis

The general strategy of variant analysis for each MH family involves: retaining all variants carried by MHS individuals and not MHN individuals; searching for variants within genes that could conceivably play a role in *RYR1* and/or ECC regulation, Ca^{2+} homeostasis or energy metabolism within myocytes; and making the assumption that there is a single unknown MH-pathogenic variant within the family.

1.6.1 Family A

In this family the male proband (II-5, Figure 1.5) had a suspected MH reaction following administration of volatile anaesthetic agents. IVCT results determined the proband and others within the family to be MHS (Figure 1.5). Subsequent analysis by DNA sequencing failed to identify a potential pathogenic variant in *RYR1*, *CACNA1S* or *STAC3*.

1.6.1.1 Approach to variant analysis

WES data was previously obtained from five individuals, two diagnosed MHS (II-2 & II-5) and three MHN (III-1, III-3 & III-4), corresponding to DNA numbers 1480, 1253, 1908, 1652 and 1653, respectively. The IVCT values for the MHS individuals are above borderline, although the proband, AII-5, had relatively low contracture values of 0.6 g and 0.4 g for halothane and caffeine, respectively. gDNA from two other individuals, one MHS (II-3) and one MHN (III-2), corresponding to DNA numbers 1479 and 1576, respectively, was also available. Variant analysis was carried out with the presumption that all MHS individuals carried the unknown pathogenic variant and all MHN individuals did not carry the pathogenic variant.

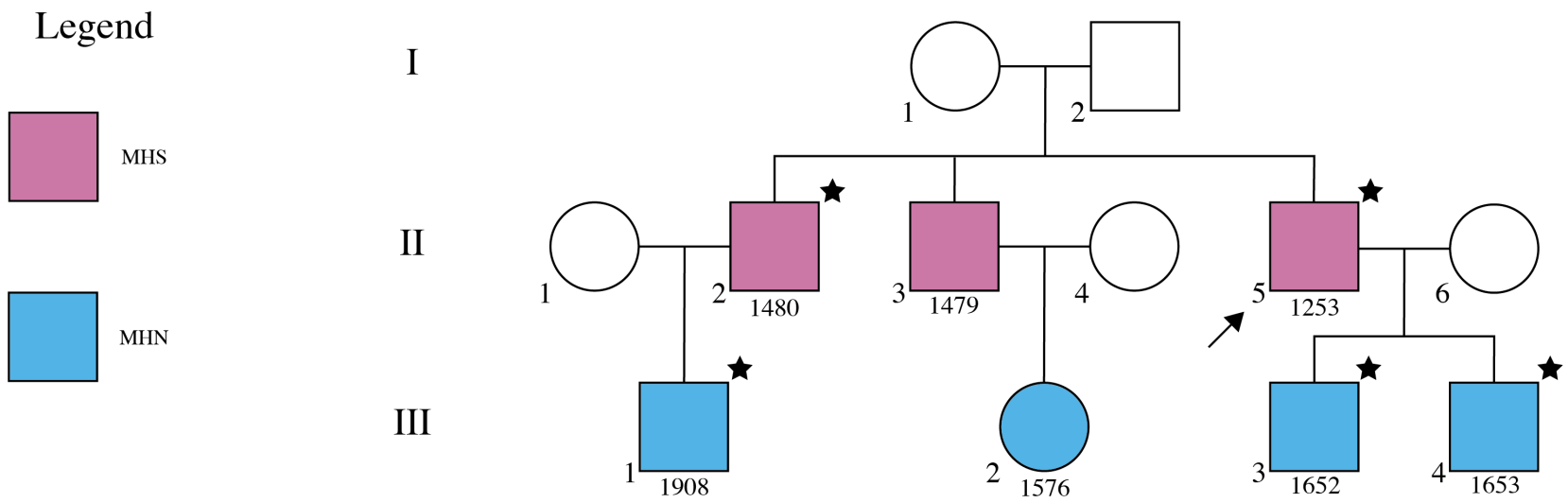

Figure 1.5: Family A pedigree

DNA numbers are given if available for each individual. WES data previously obtained is marked by a black star. The proband is depicted by the arrow.

1.6.2 Family B - A large kindred discordant for a pathogenic *RYR1* variant

This large, highly branched Māori family from the Manawatu region was the first known MH family in New Zealand and has been the subject of ongoing study. An *RYR1* variant was originally identified and subsequently demonstrated to be pathogenic for MH[130]. However, at least one other pathogenic variant is thought to be present within the family due to discordance between the familial *RYR1* variant and MHS. IVCT results and gDNA samples were available for over 120 individuals belonging to this family.

1.6.2.1 Pathogenic *RYR1* T4826I variant

Initial investigation identified an *RYR1* variant (NC_000019.10:g.38580094C>T, dbSNP: rs121918595), Thr4826Ile (T4826I), now accepted to be MH-pathogenic according to EMHG guidelines. Indeed, this variant is used as a positive control in experimental studies carried out for MH research. Nevertheless, the family is highly discordant for *RYR1* T4826I: a large number of MHS individuals do not harbour this pathogenic variant. Notably, MH-susceptibility isn't linked to the *RYR1* gene 19q13.2 locus in the discordant branches of the family[130], suggesting the cause of MH is not due to an *RYR1* variant for these individuals.

1.6.2.2 *RYR1* I3253T variant of uncertain significance

An *RYR1* variant (NM_000540.2:c.9758T>C, dbSNP: rs375626634), Ile3253Thr (I3253T) protein coding change, was previously discovered in this family. Two of the eleven family B individuals analysed through WES, 533 (VI-3, Figure 1.6) (MHS) and 1731 (VII-1, Figure 1.6) (MHS), carry this variant. It is unknown if this variant is causative of, or contributes to, an MH phenotype. It is possible that it may be an MH-modifying variant, meaning it has the potential to modify the effect of a second MH-pathogenic variant.

1.6.2.3 Approach to variant analysis

WES data was previously obtained from the gDNA of 10 individuals from this family belonging to two familial branches separated by three generations (Figure 1.6). Eight of these individuals are MHS (IV-2, DNA number 82; V-2, DNA number 1412; V-7, DNA number 1772; VI-1, DNA number 1367; VI-3,

DNA number 533; VI-5, DNA number 919; V-II, DNA number 1731), one MHS_h (V-4, DNA number 27) and one MHN (V-5, DNA number 26). However, due to uncertainty about the phenotype of individual 26, this individual was assumed to be of unknown phenotype throughout this project. One MHS individual, 82, carries *RYR1* T4826I.

Currently, little is understood of the MHS_h phenotype, with few known MHS_h individuals carrying an *RYR1* pathogenic variant. It is therefore possible that MHS_h individual 27 may or may not carry the MH pathogenic variant, and variant analysis and filtering was carried out with this in mind. Additionally, because MHS individual 82 carries *RYR1* T4826I, they may not be a carrier of a second suspected MH-pathogenic variant. However, this individual provides the familial link between both branches of the family being studied through WES analysis and so should carry a shared pathogenic variant. The other possibility is that there is a third pathogenic variant in the family and that the two branches being studied have both of these currently unidentified variants. Overall, these possibilities paint a complex picture and resulted in an approach to variant analysis that was modified from the general strategy. Individual 27 was not required to carry a specific variant for that variant to proceed to the final list variants of interest. Additionally, it was considered that individuals 1731 and 533 may not carry an unknown MH-pathogenic variant because of the presence of the *RYR1* I3253T variant with unknown significance. Because there is some uncertainty of the phenotype of individual 26, this individual was not used for segregation analysis, although processing of their WES data was undertaken.

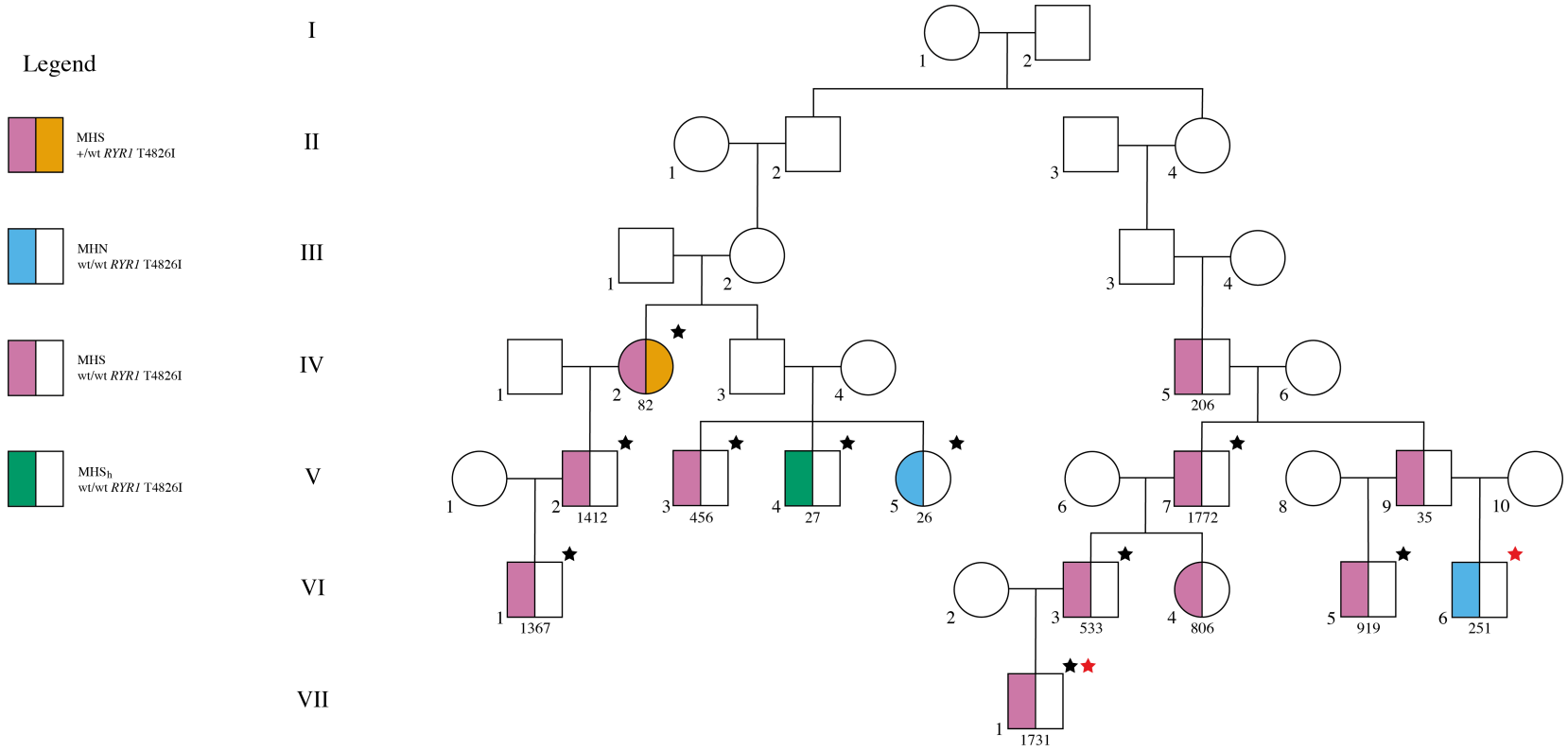

Figure 1.6: Family B pedigree

DNA numbers are given if available for each individual. WES data previously obtained is marked by a black star. WES data obtained during this project is marked by a red star. Note that there is some uncertainty of the phenotype of individual 26.

1.6.3 Family C - MH susceptibility with central core disease

In addition to MH-susceptibility, some individuals in this family suffer from CCD (Figure 1.7). CCD phenotypic data was unavailable for most individuals and is not given in the pedigree. One individual, III-7, is MHS_h.

1.6.3.1 Approach to variant analysis

WES data was previously obtained from five individuals within the family, two MHS (II-1, DNA number 1360; II-12, DNA number 1273) and three MHN (II-4, DNA number 1928; III-2, DNA number 1474; III-4, DNA number 2338). CCD is often associated with MH-susceptibility as judged by IVCT (see Section 1.2.4). The fact that MH-susceptibility may be secondary to the CCD phenotype in this family broadens the spectrum of genes that may harbour a pathogenic variant and therefore variant analysis was carried out in a more lenient manner.

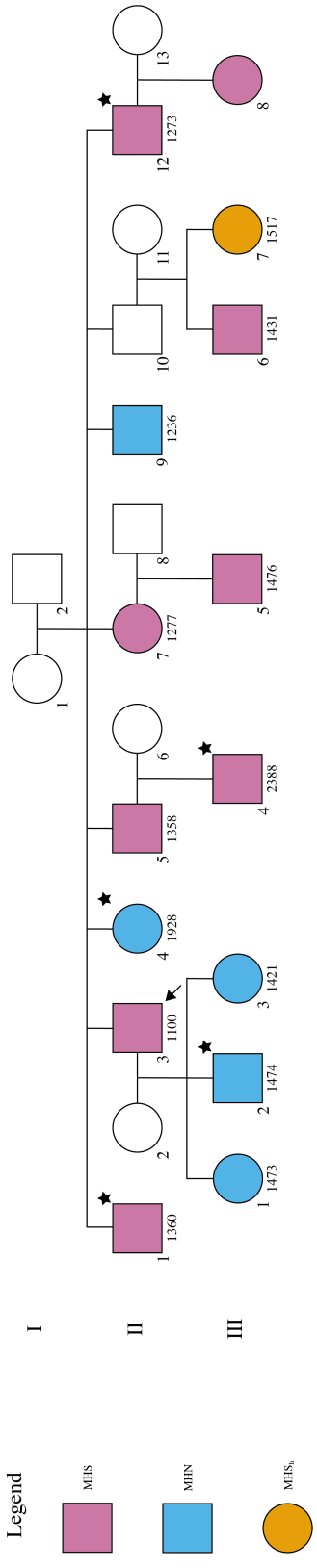

Figure 1.7: Family C pedigree
DNA numbers are given if available for each individual. WES data previously obtained is marked by a black star. The proband is depicted by the arrow.

1.6.4 Family D - discordant for a *CACNA1S* variant

A male proband (II-2, DNA number 1128) experienced an MH-like episode following administration of general anaesthetics and was later diagnosed MHS by IVCT. The IVCT determined a number of others in the family to be MHS or MHS_c (Figure 1.8).

1.6.4.1 *CACNA1S* T1009K variant

Previous WES analysis identified a *CACNA1S* variant, NM_000069:c.3026C>A (dbSNP: rs200224590) with a predicted coding change threonine to lysine at amino acid position 1009 (T1009K), which was found to have a degree of co-segregation with MHS in the family (Figure 1.8). Noteably there is an individual, III-6, discordant for the variant for whom IVCT values are significantly above borderline (1.9 g for halothane, 1.3 g for caffeine).

So far, the significance of this *CACNA1S* variant is unclear. The variant has been found in a second MH-susceptible family for which no pathogenic *RYR1* variant has been identified[51,52]. It is predicted to be pathogenic by multiple prediction algorithms (CADD score: 29.8), both due to its presence in a highly conserved pore domain region of DHPR and the nature of the amino acid substitution: An uncharged polar residue to a basic residue. However, due to current limitations of experimental systems exploring functional consequences of variants that do not fall within *RYR1*, functional studies have not yet been reported for *CACNA1S* T1009K. Overall, the variant is predicted to be likely pathogenic using the American College of Medical Genetics and Genomics standard and guidelines[131].

1.6.4.2 Involvement of an unknown variant

The presence of individual III-6 discordant for *CACNA1S* T1009K suggests that another as yet unknown MH-pathogenic variant may be present within the family.

1.6.4.3 Approach to variant analysis

WES data was previously obtained from gDNA of four individuals including the proband (II-2, DNA number 1128), two other MHS individuals (III-1, DNA number BW1; III-3, DNA number 1172) and an MHN individual (III-2, DNA

number 932). WES data from the discordant family member, III-6, was not obtained. The presence of the *CACNA1S* T1009K variant raises the possibility that an unknown pathogenic variant may not be carried by all MHS individuals within the family, and consequently, variant analysis was modified accordingly.

It is possible that there is a variant pathogenic for MH in individual III-6, who does not harbour the *CACNA1S* T1009K variant, and which is not carried by other MHS individuals within the family. This would assume *CACNA1S* T1009K is pathogenic for MH. As genotype filtering was carried out under the assumption that all three MHS individuals, for which WES were obtained, harbour the unknown MH-pathogenic variant, and WES data from individual III-6 was not obtained, the variant will not be identified if this is the case. The halothane contracture force recorded for individual III-3 (DNA number 1172), 4.2 g, is much higher than other MH individuals within this family. If *CACNA1S* T1009K is pathogenic, and there is another MH-pathogenic variant within the family, the cause of this extreme halothane contracture value may be due to this individual harbouring both MH-pathogenic variants. Nevertheless, WES data from individual 1172 was not studied alone due to time constraints and low WES data quality (see Chapter 3.2.2).

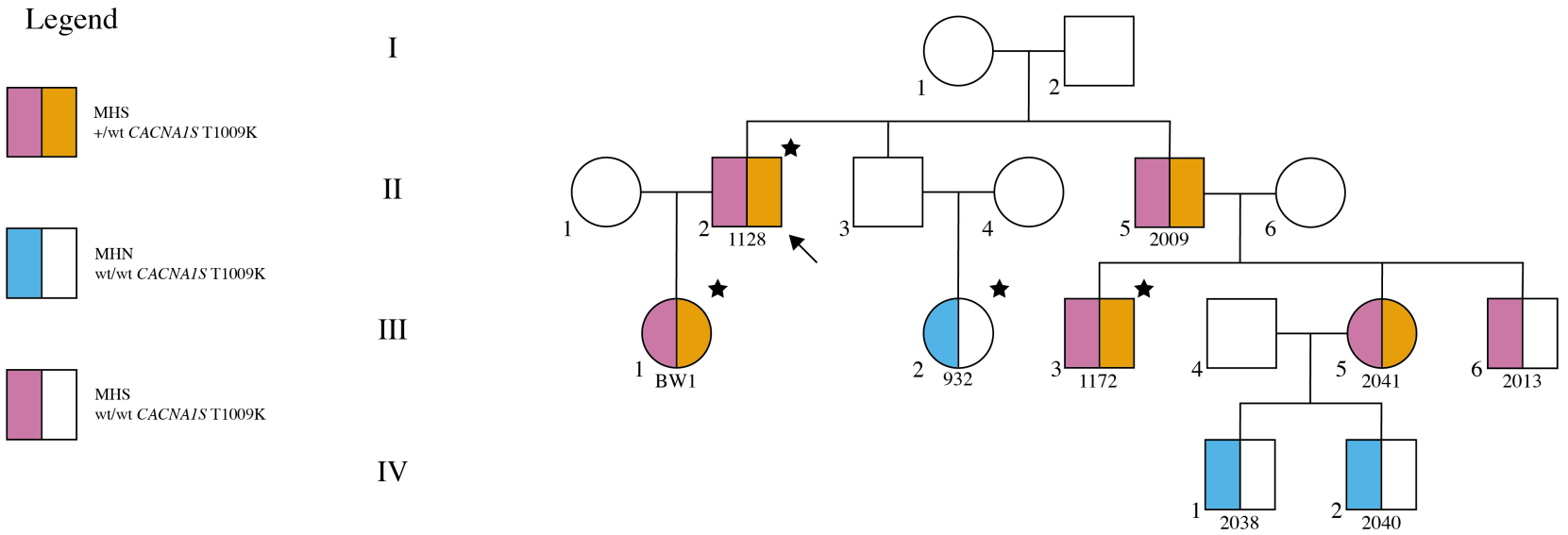

Figure 1.8: Family D pedigree

DNA numbers are given if available for each individual. WES data previously obtained is marked by a black star. The proband is depicted by the arrow.

1.6.5 Family E - MH susceptibility and a myopathy

An MH-like episode was observed in the male proband (III-7, Figure 1.9) following administration of volatile anaesthetic agents for acute appendicitis and a subsequent IVCT diagnosed the patient MHS_{hc}. The patient and other family members (for which complete phenotypic data is not available) had been previously diagnosed with mild myotonic dystrophy type I (DM1); this was associated with a *DMPK* gene CTG repeat expansion (87 repeats) present within the family[132,133]. Other family members have since been tested by IVCT, with both MHS_{hc} and MHS_{h} diagnoses in addition to MHN (Figure 1.9). The branch of the family fathered by II-2 resides in Australia and thus DNA availability was limited for these individuals.

1.6.5.1 *RYR1* R1679H variant

An *RYR1* variant (NM_000540.2:c.5036G>A, dbSNP: rs146504767), with predicted coding change arginine to histidine at amino acid position 1679 (R1679H), was identified in the family (Figure 1.9). The variant has a strong degree of segregation with MHS, however, it was disproven to be solely causative due to substantial discordance: individuals IV-1 and IV-4, MHS_h and MHS_{hc}, respectively, did not carry R1679H, whilst MHN III-5 did carry the variant. Nevertheless, *in vitro* calcium release experiments using HEK293T cells transiently expressing the variant revealed that it did confer a small yet significant increase in sensitivity to *RYR1*-channel agonist 4-C*m*C compared to cells expressing WT *RYR1* (personal communication, Professor K M Stowell, SFS, Massey University). Collectively, these results indicate that the *RYR1* R1679H variant may contribute to the presence of MH in this family, but it does not wholly explain it due to the discordance between the variant and MH-susceptibility observed.

1.6.5.2 *DMPK* trinucletoide expansion

A *DMPK* CTG repeat expansion of greater than 34 is considered to be pathogenic for DM1[132]. Pathogenic CTG repeat expansions were identified in the proband of family E as well as other family members (Figure 1.9). Genetic linkage exists between the repeat and the *RYR1* R1679H variant due to the proximal location of the genes on chromosome 19: *DMPK* is located at chromosomal position 19q13.32, *RYR1* at position 19q13.2, and so it is therefore not surprising

to see segregation of the *DMPK* repeat and the *RYR1* variant for all individuals tested within the family.

1.6.5.3 Involvement of an unknown variant

DM1 is thought to have the potential to lead to abnormalities in Ca^{2+} homeostasis due to reported alternative splicing of *RYR1* and *ATP2A1* [133]. It is tempting to presume that this repeat variant, along with the putative disease-modifying *RYR1* variant, R1679H, explains the presence of MH in this family. However, the presence of the CTG expansion and the *RYR1* variant is not completely concordant with susceptibility to MH, with MHN individual III-5 carrying 69 *DMPK* CTG repeats and the *RYR1* R1679H variant. Additionally, MHS_{hc} individual IV-4 carries neither the *DMPK* CTG expansions or *RYR1* variant. Overall, it seems possible that the presence of a so far unknown variant may be the cause of MHS in the family, with the potential that other variants, such as the *RYR1* R1679H and *DMPK* repeat expansion, have a modifying effect.

1.6.5.4 Approach to variant analysis

WES data was previously obtained from individuals II-2 (MHS), III-8 (MHS, DNA number 1247), IV-2 (MHS, DNA number MS1) and IV-6 (MHN, DNA number 1680). It is important to note that although individual IV-2 has been diagnosed MHS_{hc} through IVCT, the contracture values, 0.2 g and 0.3 g for halothane and caffeine, respectively, are borderline, meaning a diagnosis of MHS_c may have been made in another IVCT testing facility. Conversely, individual IV-6 was diagnosed MHN although the IVCT contracture value for halothane was 0.2 g (0 g for caffeine), a level which may have constituted a diagnosis of MHS_h in another IVCT testing facility. Additionally, individual II-4 was diagnosed MHN with contracture values of 0.2 g for halothane and 0.4 g for caffeine, certainly classifiable as borderline. This is significant as it raises the possibility of an MH-pathogenic variant being inherited maternally on the branch of the family containing the proband. This would suggest the presence of a second unknown MH-pathogenic variant on the other branch. Further WES were obtained from individuals IV-4 (MHS, DNA number 44), III-7 (MHS, DNA number 2253), III-9 (MHS, DNA number 1379) and IV-7 (MHN, DNA number 2259).

Due to the complexity of family E's phenotype and the variability of IVCT results between individuals, variant genotype filtering (see Chapter 3.3.1) was

adapted accordingly. For example, individual IV-2 (DNA number MS1) was not required to carry a variant for the variant to proceed through the genotype filter due the individual's borderline IVCT values.

1.6.6 Family F

WES data was obtained from the gDNA of three MHS individuals (4003, 4004, 4005) from this Australian family. WES from MHN family members were not available which eliminated the possibility of filtering variants by genotype with the bioinformatic pipeline. Additionally, gDNA from other individuals within the family was not available, limiting the opportunity to carry out segregation analysis for any variants of interest. This meant that a higher burden-of-proof was required to pursue a variant of interest due to the difficulty in obtaining additional gDNA samples. Overall, these two issues culminated in a challenging variant analysis task when compared to other families. A pedigree for this family was not available.

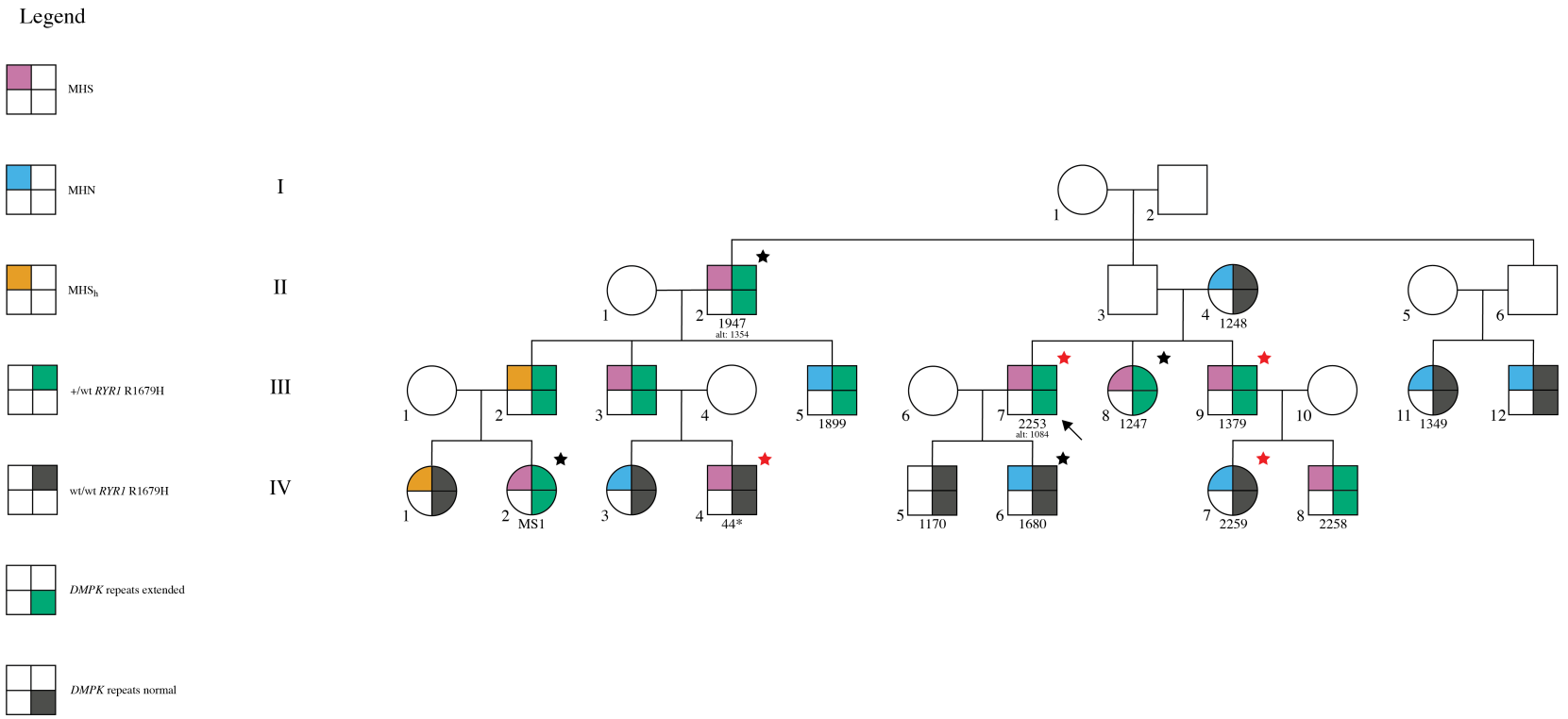

Figure 1.9: Family E pedigree

DNA numbers are given if available for each individual. IV-4, with DNA number 44 marked by the asterisk, is a placeholder DNA number as a DNA number was not available. WES data previously obtained is marked by a black star. WES data obtained during this project is marked by a red star. The proband is depicted by the arrow. 'alt' refers to alternative DNA numbers.

Chapter 2

General Methods

2.1 Materials

- 5x HOT FIREPol® Blend Master Mix Ready to Load with 7.5 mM $MgCl_2$ (Cat. No. 04-25-00120), Solis BioDyne, Tartu, Estonia
- Gel Loading Dye, Purple (6X) (Cat. No. B7024S), NEB®, Ipswich, MA, USA
- Invitrogen™ 1 Kb Plus DNA Ladder (Cat. No. 10787018), ThermoFisher Scientific
- Wizard® Genomic DNA Purification Kit (Cat. No. A1120), Promega, Madison, WI, USA
- Gentegra®-DNA (Cat. No. GTD4010-PBC), Gentegra®, Pleasanton, CA, USA
- LightCycler® 480 High Resolution Melting Master (Cat. No. 04909631001), Roche Diagnostics, Mannheim, Germany
- PerfeCTa® qPCR ToughMix® (Cat. No. 97065-958), QuantaBio, Beverly, MA, USA
- FastStart™ Taq DNA Polymerase GC-RICH Solution 5x concentrated (Cat. No. 12032902001), Roche Diagnostics, Mannheim, Germany
- LightCycler® 480 Genotyping Master (Cat. No. 04707524001), Roche Diagnostics, Mannheim, Germany

- ScrFI restriction enzyme (Cat. No. R0110S), NEB®, Ipswich, MA, USA
- 10x CutSmart® Buffer (Cat No. B7204S), NEB®, Ipswich, MA, USA
- LightCycler® 480 Multiwell Plate 96, white (Cat. No. 04729692001), Roche Diagnostics, Mannheim, Germany
- ExoSAP-IT™ PCR Product Cleanup Reagent (Cat. No. 78200.200.UL), Applied Biosystems™, ThermoFisher Scientific

2.2 Ethical approval

The Central Regional Ethics Committee and Ministry of Health (New Zealand) ethics approval code is MWH/03/04/018. Patient approval was given to analyse patient genetic data for MH-related research. Incidental findings were not reported nor analysed.

2.3 Sterile water

Unless otherwise stated, all dH_2O was Milli-Q® filtered.

2.4 PCR amplification

Unless otherwise stated, all PCR reactions were carried out in the following manner. A reaction mixture containing 4 μL 5x HOT FIREPol® Blend Master Mix Ready to Load with 7.5 mM $MgCl_2$, 1 μL each of 5 μM forward and reverse primer, 14.5 μL dH_2O, and 0.5 μL 50 ng μL^{-1} gDNA was set-up on ice in a 0.2 mL PCR-tube. A thermocycler machine was used with the following conditions: 12 minutes 95 ℃; 30 cycles of 95 ℃ for 10 seconds, variable primer annealing temperature for 30 seconds, 72 ℃ for 20 seconds; followed by 72 ℃ for 5 minutes.

2.5 DNA agarose gel electrophoresis

Unless otherwise stated, DNA fragments were separated using 0.8% agarose (for fragments > 0.3 kb) or 1% agarose (for fragments < 0.3 kb), TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) and 0.2 μg mL^{-1} ethidium bromide. Six x DNA Gel Loading Dye (NEB®, Cat. No. B7024S) was used, if the DNA did not already contain HOT FIREPol®Blend Master Mix Ready to Load, to assist with DNA loading and to visualise DNA migration. Approximately 200 ng of DNA sample was loaded into each lane alongside 3 - 5 μL of 50 ng μL^{-1} 1 Kb Plus DNA ladder (Invitrogen^{TM}). Electrophoresis was carried out in TAE buffer at 100 V for 40 - 60 minutes in TAE and the resulting DNA bands visualised using ultraviolet light.

2.6 DNA concentration measurements

The DeNovix DS-11 FX+ spectrophotometer (herein referred to as DeNovix) was used to measure DNA concentrations by absorbance at 260 nm for both ds-DNA and ss-DNA. Alternatively, DNA quantification was estimated by carrying out agarose gel electrophoresis on DNA alongside a 50 ng μL^{-1} 1 Kb Plus DNA ladder (Invitrogen^{TM}), which contains DNA fragments of known concentrations, to aid in quantification.

2.7 Whole exome sequencing

2.7.1 Sample and library preparation

gDNA was extracted from patient muscle tissue or from whole blood.

2.7.1.1 Muscle tissue gDNA extraction

Approximately 150 mg of patient muscle tissue was obtained and frozen in liquid nitrogen. The frozen tissue was ground into a powder using a steel mortar. Six hundred μL nuclei lysis solution (Wizard® Genomic DNA Purification Kit) was added to the powder and the solution was mixed with a glass homogeniser followed by a 15 minute incubation at 65 °C. The solution was cooled to room temperature and 200 μL protein precipitation solution (Wizard® Genomic DNA

Purification Kit) was added to the solution and mixed using a vortex. It was cooled on ice for five minutes and centrifuged at 9500 x g for four minutes. The supernatant was transferred to a tube containing 600 μL isopropanol and mixed gently followed by centrifugation at 9500 x g one minute. The supernatant was removed and 100 μL of 70% ethanol was added followed again by centrifugation at 9500 x g for one minute. The supernatant was removed and, following air drying of the DNA pellet, 100 μL TE buffer was added for resolubilisation of the DNA.

2.7.1.2 Whole blood gDNA extraction

gDNA was isolated from 3 mL patient whole venous blood using the Wizard® Genomic DNA Purification Kit kit (Promega) according to the manufacturer's instructions.

2.7.1.3 gDNA quality control

The quality of extracted gDNA was assessed by two methods: DNA gel electrophoresis and the DeNovix spectrophotometer. For DNA gel electrophoresis, approximately 200 ng of gDNA was run on an agarose gel and analysed for migration distance, mass and degradation. For the DeNovix spectrophotometer, 1 μL of gDNA was loaded and A_{260}/A_{280} ratio readings were obtained, with a ratio of 1.7-2.0 indicating suitable gDNA quality.

2.7.1.4 gDNA preparation for shipment

For gDNA samples sent for MGI-sequencing (2019 samples), GenTegra®-DNA was used for ambient temperature storage and transport of gDNA samples. gDNA quality was analysed using the DeNovix spectrophotometer and by DNA gel electrophoresis.

gDNA was diluted as necessary to a concentration of 50 ng μL^{-1} with TE buffer. Following the manufacturer's instruction, 20 μL of diluted DNA was pipetted into the GenTegra®-DNA 0.5 mL tubes and mixed by pipetting to achieve a final mass of 1 μg gDNA per sample. gDNA was dried using a Savant SC100 SpeedVac machine.

2.8 Genotyping methods

2.8.1 Sanger sequencing

To confirm the presence of a variant, Sanger sequencing was carried out through the Massey Genome Service on 400 ng of purified (using ExoSAP-IT™ PCR Product Cleanup Reagent (Applied Biosystems™) according to the manufacturer's instructions) PCR products using 4 pmol of primer (see Appendix A.1 for primer sequences). Geneious® version 9.1.8 was used for computational analysis of sequencing chromatrograms.

To confirm the presence of a variant, Sanger sequencing was carried out on 200 ng of PCR product with 4 pmol μL^{-1} primer through the Massey Genome Service using BigDye™ Terminator Version 3.1 chemistry on an ABI3730 DNA Analyzer.

2.8.2 High resolution melting analysis

To perform high-throughput genotyping, high resolution melting (HRM) analysis was carried out. HRM analysis is a rapid and relatively inexpensive technique which exploits differences in DNA melting behaviour of small amplicons which differ by only a single base pair.

2.8.2.1 Principle of HRM

Fluorescent, saturating DNA binding dye is present within the HRM reaction mixture to monitor DNA melting over a temperature range. PCR is initially carried out to produce an amplicon (up to ∼100 bases) containing the position of the SNV of interest. Melting curve analysis is then carried out. Firstly, the temperature is increased for denaturation of the amplicon to occur, and then the temperature is reduced so as to encourage the single strands to anneal (duplex formation) even if there is a single nucleotide mismatch between two strands (which occurs if there is a SNV present in a heterozygous state). Duplexes containing a mismatch are referred to as heteroduplexes. For a heterozygous variant, there will be populations of both homo- and hetero-duplexes present during melting curve analysis. These two populations are detected as the temperature is increased as they melt at different rates, therefore leading to differential fluorescence changes when compared to reactions containing only homoduplexes. Homozygous SNVs

can also be differentiated from wild-type amplicons as these homoduplexes will melt at different rates. The results of melting curve analysis are presented in the form of a graph with temperature on the x-axis and change in fluorescence (465-510 nm)/time on the y-axis. A sample with a SNV will have a characteristic heteroduplex double peak compared to a wild-type sample with a characteristic homoduplex single peak. A sample with a homozygous SNV in a homozygous state will have a characteristic shifted homoduplex single peak.

2.8.2.2 HRM assay method

HRM assays were carried out on a LightCycler® 480 Instrument (Roche). This machine performs PCR followed by melting curve analysis of the produced amplicon. Unless otherwise stated, HRM reactions were carried out in the following manner: Individual reactions containing 5 μL LightCycler® 480 High Resolution Melting Master (Roche), 0.6 μL each of 5 μM forward and reverse primer, 1.2 μL 25 mM $MgCl_2$, 2.1 μL PCR-grade H_2O and 0.5 μL of 50 ng μL^{-1} gDNA or 0.5 μL PCR-grade H_2O as a negative control. Each 10 μL reaction was pipetted into a single well of a LightCycler® 480 Multiwell Plate 96 (Roche). The following conditions were carried out on the LightCycler® 480 Instrument II: 10 minutes 95 ℃ (initial denaturation); 45 cycles of 95 ℃ for 10 seconds (denaturation), variable primer annealing temperature (annealing) for 10 seconds, 72 ℃ for four seconds (amplification); and lastly for melting curve analysis, 95 ℃ (denaturation), 40 ℃ for 1 minute (amplicon annealing), and 76 ℃ for one second at a ramp rate of 0.02 ℃ per second to 92 ℃, with 25 fluorescence acquisitions taken for each ℃ increase. Control genotypes confirmed by Sanger sequencing were assayed alongside unknown genotypes.

2.9 Bioinformatic pipeline

All sequence read data was obtained in the form of trimmed (removal of adaptor sequences required for sequencing) and de-multiplexed (removal of sample-specific adaptor sequences) paired fastq files from the WES service provider. Additionally, all read data obtained had been quality filtered according to the specifications of the company.

All commands used in this project are given in the 'scripts' directory of the following GitHub repository: https://github.com/sormond/Masters.

Supplementary material for this thesis is given in the 'supp_material' directory of the above GitHub repository.

2.9.1 Alignments

Fastq files were aligned to the hg38 human reference genome using bwa[92] mem (version 0.7.17) to create sam alignment files (scripts/bwa_samtools.sh)

Samtools[96] (version 1.7) view, sort and index functions were used to convert sam files into sorted bam files with an index file for faster computation (scripts/bwa_samtools.sh).

2.9.2 Marking PCR duplicates

Picard's[134] (version 2.8.1) MarkDuplicates was used to mark PCR duplicates within the bam file (scripts/mark_duplicates.sh).

This step was not initially a part of the pipeline but was added for the new 2019 MGI-sequenced WES for family B and E. The step was also carried out on bam files for samples to be used for copy-number variant (CNV) detection prior to CNV detection as this was required for the programme (see Section 2.9.8 and Chapter 3.5). All WES samples within final family VCFs of families B and E were obtained from bam files for which duplicates were marked, whereas WES samples within all other family VCFs were not obtained from bam files for which duplicates had been marked. For family B and E, due to the presence of MGI-sequenced samples, the READ_NAME_REGEX flag of the MarkDuplicates tool was set to null to avoid optical duplicate detection, which is only applicable to Illumina-sequenced samples. This flag was applied to both MGI-sequenced and Illumina-sequenced samples from these families for consistency.

2.9.3 Variant calling

The Genome Analysis Toolkit (GATK)[135] (version 4.1.3.0) was used to call variants. This process requires three tools: HaplotypeCaller, CombineGVCFs and GenotypeGVCFs. HaplotypeCaller is used to create an initial intermediate variant file, a gVCF, for each individual (scripts/variant_calling_gatk.sh). Then, CombineGVCFs is used on these individual gVCFs to create a whole-family gVCF. Lastly, joint-genotyping is carried out on each family with GenotypeGVCFs.

2.9.4 Variant annotation

SnpEff[106] version 4.3t was used for variant categorisation annotation (scripts/snpeff.sh). This annotation step was carried out as the subsequent ANNOVAR annotation did not apply 'Refseq' accessions to the variants, which was necessary for the addition of GO annotations (see Section 2.9.4).

VarAft[123] version 2.16, also described in Chapter 3.3, was used to apply ANNOVAR[104] and other functional effect annotations. VarAFT varied from other programmes used in this work in that it is has a graphical user interface as opposed to command line interface and therefore commands used for each step cannot be given. However, an explanation of how the programme was used can be found below. A list of all annotations and the database versions from which annotations were applied by VarAFT can be found in Appendix A.32.

2.9.4.1 VarAFT

VarAFT takes a VCF as input. For family VCFs, 'Multi Samples' is specified within VarAFT. Additionally, the genome version, hg38, and the dbSNP version to be used (147 for all annotations in this project), are specified. VarAFT then carries out annotation of the VCF. An annotation file, which differs significantly from VCF format, is produced which contains all variants and their respective annotations.

To filter variants in VarAFT based on their annotations and genotype of a sample, the manual filtering tool is selected, followed by 'Custom Analysis'. The desired genotype is selected for each sample. For all families in this study, all sample genotypes were set to 'ND' (not defined) because genotype filtering was carried out using a custom python script (described in Chapter 3.3.1) prior to VarAFT annotation and filtering. Variants are then filtered by variant type and AF by unchecking the necessary boxes. These variants of interest are then exported as a txt file. These files are referred to as 'variants of interest' lists.

2.9.4.2 addAnnotations python script

A custom python script was written to apply the snpEff-applied RefSeq annotations from the VCF which were not carried on to the VarAFT output variant lists (scripts/ addAnnotations.py). This was necessary for the application of GO annotations.

2.9.4.3 GO annotations

Gene ontology (GO) terms were applied to the variants of interest lists using a custom R script (scripts/addGOanno.R).

2.9.5 Querying of variants against gene lists

A custom python script was used to query VarAFT output variants of interest lists against genes of interest lists (scripts/geneList.py). Gene lists are available at https://github.com/sormond/Masters/supp_material/.

2.9.6 Genotype filtering using gtFilter

A custom python programme, gtFilter (scripts/gtfilter), was used to filter variants based on variant genotype and sample phenotype. The programme was used according to the instructions within the README file [†]. The script was run on the snpEff-annoted family VCFs prior to VarAFT. A 'pos' and 'neg' text file was given for each family according to each family's specific considerations (see Chapter 1.6 & Chapter 3.3.1). These files can be found in supplementary material. The gtFilter programme was not used on the family F family VCF as there was no MHN WES sample from this family.

2.9.7 Family F snpSift

SnpSift [106] (version 4.3t, part of the snpEff programme) was used to pre-filter the family F VCF prior to VarAFT annotation and filtering (scripts/snpsift_familyF), for reasons described in Chapter 3.6.

2.9.8 Detection of copy-number variation

CNVkit [126] (version 0.9.7.dev0) was used for detection of copy number variants (CNV) (scripts/cnvkit.sh). See Chapter 3.5 for details of how detection of CNV was carried out for each family.

[†]README file available at https://github.com/sormond/gtfilter

2.9.9 Coverage analysis

2.9.9.1 WES kit target coverage

Picard's CollectHsMetrics was used for calculation of coverage over the WES kit target region (scripts/collectHsMetrics.sh). This tool accepts the bam file, target and bait intervals, as well as the reference genome fasta file used for the alignment. Target and bait intervals are created using Picard's BedToIntervalList. Samtools faidx is first used to create a dictionary file of the genome reference which is required for the BedToIntervalList tool. The target interval is the genomic coordinates of the kit's target sequencing region, whilst the bait interval is the genomic coordinates of the region that the kit's baits are designed to anneal to. The bait design was not available for the WES kits used for this project. Therefore, the target region interval file was given for both bait and target inputs.

Plots of average target coverage and proportion of target bases with low coverage were made using the ggplot2[136] R package version 3.2.1 (scripts/coverage_plots.R). Prior to running this script, the CollectHsMetrics output was converted into a csv file and a 'Year' column was added containing the year that the WES sample was sequenced. A csv file containing coverage data for all families is available at:
https://github.com/sormond/Masters/supp_material/coverage_supp_material.csv.

2.9.9.2 Coverage over a specific region for CNV analysis

Mosdepth[137] version 0.2.3 was used to calculate per-base coverage bed files over the WES kit target region (scripts/mosdepth.sh). The Sushi[138] version 1.22.0 R package was used to visualise coverage of the genomic region containing a suspected CNV (scripts/sushi_plots.R).

2.9.10 Visualisation of alignments

Tablet[139] version 1.19.05.28 was used to visualise reads aligned to the reference genome.

Chapter 3

Bioinformatic methods and analyses

3.1 Introduction

This chapter will describe bioinformatic method development and the various analyses carried out to identify variants of interest (VOI) from each family. The VOI identified using these methods are presented and discussed in Chapter 4. Figure 3.4 gives the workflow of the bioinformatic pipeline.

Throughout the project, methods for identification of VOI were altered and refined over time. For example, problems with whole exome sequencing (WES) quality and compatibility of WES samples were identified early on, and it was determined that the software used for variant filtering could not handle these issues, possibly leading to a proportion of variants of interest being inadvertently discarded. Due to this, a python programme for genotype filtering was developed and successfully employed for the remainder of the project, and WES was carried out on other individuals from two of the families.

This chapter will begin by discussing measurements of read coverage and the identification of coverage issues in some WES samples. Then, it will discuss the development, logic and use of the methods and programmes used for variant annotation, variant filtering, and selection of variants of interest. The general methods and scripts used for the bioinformatic steps can be found in Chapter 2.

3.2 Coverage analysis

There are a multitude of tools available for measurements of sequencing coverage, but for proper analysis of targeted sequencing data such as WES, these tools need to calculate the coverage over the genomic coordinates of the target regions covered by the WES kit (WES kits vary in their exome target regions). Coverage analysis provides a good estimate of the quality of WES data: higher coverage of the target region generally correlates with increased true positive variant identification[80].

3.2.1 Methods of coverage analysis

Target regions are given by the kit manufacturers in the form of a *bed* file. At minimum, a bed file contains three columns in the following order: the chromosome name; the beginning base coordinate; and the ending base coordinate. A bed file may have additional columns containing other information, such as the genes that fall within a given coordinate, or the read depth at a given coordinate.

For the WES samples in this study, multiple WES kits were used: SureSelect^{XT} Human All Exon V6 (Cat. No. 5190-8863, Agilent); Nimblegen SeqCap v3 (Cat. No. 06465684001, Roche Nimblegen); and a SureSelect Human All Exon (Agilent) kit produced during or prior to 2010 (the catalogue number or version number for this kit could not be confirmed). The target regions differ among these kits. The target region bed file for the SureSelect Human All Exon (Agilent) kit could not be obtained for family B WES sample 533, so the coverage metrics were calculated using the SureSelect^{XT} Human All Exon V6. A list of WES samples and the WES kits used is available in Appendix A.33.

Because these target region bed files were given in hg37 coordinates, to allow compatibility with the genome version used throughout the pipeline, target region bed files were converted from human reference genome hg37 coordinates to hg38 coordinates with the Galaxy LiftOver online tool using the public server at usegalaxy.org on the Galaxy web platform[140].

Picard's[134] CollectHsMetrics tool was used for calculation of coverage metrics over the WES kit target region (scripts/collectHsMetrics.bsh). Additionally, mosdepth[137] was used for coverage analysis of more specific genomic intervals. This is described further in section 3.5.

3.2.2 Results of coverage analysis

Data of coverage results obtained using CollectHsMetrics for all WES samples are available in the supplementary material directory of https://github.com/sormond/Masters (supp_material/coverage_supp_material.csv). Plots of average target coverage (the mean coverage over the genomic region targeted for capture by the WES kit) and the proportion of target bases with low coverage are given in Figures 3.1 and 3.2. Average target coverage and proportion of target bases with low coverage are generally inversely correlated, however this is dependent on evenness of coverage so it is still useful to plot and assess both of these metrics.

The results suggest that there is a large degree of variance of WES quality for different WES samples within each family. In these plots, the WES samples are grouped, using colour sets, by the year in which the WES samples were sequenced. WES obtained more recently appear to be of higher quality. One anomaly to this trend is WES sample 1908 from family A, which was sequenced in 2017, the year following sequencing of the other WES samples from this family. As the WES kit and platform are the same for all samples in this family, it is possible that low gDNA quality for this sample was the cause of lower WES quality.

Importantly, the coverage results indicate that for most families, there are a number of WES samples in which the WES quality may be problematic. For WES sample 533 from family B, the average target depth is significantly lower than all other WES samples within this family, and almost 50% of target region bases have less than 10x read depth. This may be due to the fact that the WES kit target region bed file used for the analysis differed from the actual WES target kit used: it is likely that the more recent WES kit (SureSelect^{XT} Human All Exon V6) has a larger target region due to improved predictions and confirmation of exonic regions of the genome. Prior to WES of a new batch of samples (2019 samples) from family E, approximately 60% of target region bases had less than 10x read depth for all WES samples. WES sample 1928 from family C had significantly lower average target coverage compared to WES samples sequenced in the same batch, possibly due to low gDNA quality.

3.2.3 Family B WES kit differences

For family B, WES have been obtained from individuals on two relatively distant branches of the family (see Figure 1.6). The WES samples from one

branch, corresponding to individuals 26, 27, 82, 456, 1367 and 1412, were sequenced recently (2016 and 2017) with the same WES kit (SureSelect^{XT} Human All Exon V6). The WES samples from the other branch (excluding WES from 2019), corresponding to individuals 533, 919, 1731 and 1772, were sequenced in 2010 and 2012 with WES kits that differed from the newer SureSelect^{XT} Human All Exon V6 WES kit samples. This has implications for identification of variants shared amongst these two branches of the family: variants may fall within regions which are not targeted by both of the different WES kits used, and so will be identified in only one branch and thus may be excluded during variant filtering. To measure the extent of this problem, the percentage of bases where coverage was low (defined as a read depth of less than 15) for at least one of the 10 family B WES samples (excluding WES from 2019) over the most recent WES kit release, SureSelect^{XT} Human All Exon V6, was calculated (scripts/low_cov_test.bsh). This script used mosdepth for calculations of coverage and Bedtools[141] version 2.29.0 merge and intersect tools. This determined that coverage was low in 57% (3.5 x 10^7 bases/ 6.1 x 10^7 bases) of the total number of bases within the SureSelect^{XT} Human All Exon V6 target region for at least one of the 10 WES samples. Recognition of this issue was important for informing how genotype variant filtering was carried out, as discussed in Section 3.3.1.

3.2.4 New WES for family B and E

Due to problems with WES kit differences in family B, and low coverage in family E, further WES was carried out (2019 samples). Ideally, re-sequencing of family D and family F would have been carried out due to low coverage, but this was not possible due to financial constraints.

All new WES was carried out using DNA nanoball sequencing technology on an MGISEQ-2000 machine with the SureSelect^{XT} Human All Exon V6 kit. For family B, MHS individual 1731 was re-sequenced (1731_new), and MHN individual 251 was sequenced for the first time. Both of these individuals belong to the branch of family B containing older WES samples, and so the addition of these new WES was intended to enhance variant detection from that side of the family. For family E, MHS individuals 44, 1379 and 2253, and MHN individual 2259, were sequenced for the first time. The addition of discordant individual 44, the only MHS individual in family E that does not carry the *RYR1* R1679H variant or the extended *DMPK* repeat allele, was useful as the MH-susceptibility in this

individual cannot be explained by the presence of these two known variants. The addition of the family E proband 2253 was useful as this individual is *bona fide* MHS, having previously suffered an MH episode.

Methods for preparation of gDNA for WES are described in Chapter 2.7.1. Agarose gel electrophoresis images of gDNA samples for these new WES from family B and E can be found in Appendix A.1.

The coverage of the new WES samples was relatively high (2019 samples from Figure 3.1) and this new WES data was used going forward. Herein, the new WES sequencing samples will be referred to as 'MGI-sequenced WES samples' or '2019 WES samples'.

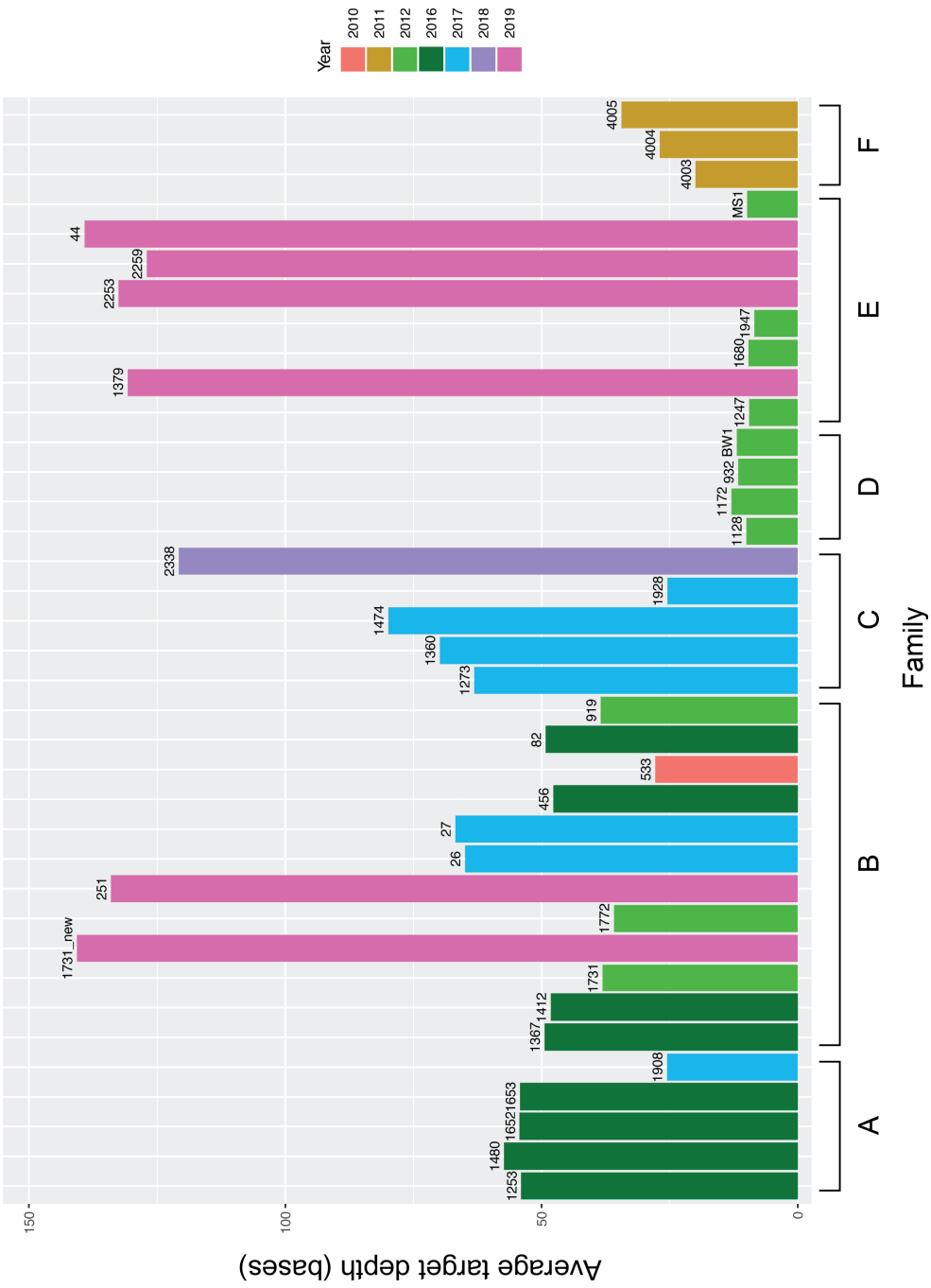

Figure 3.1: Average coverage over the WES kit target region for all WES samples

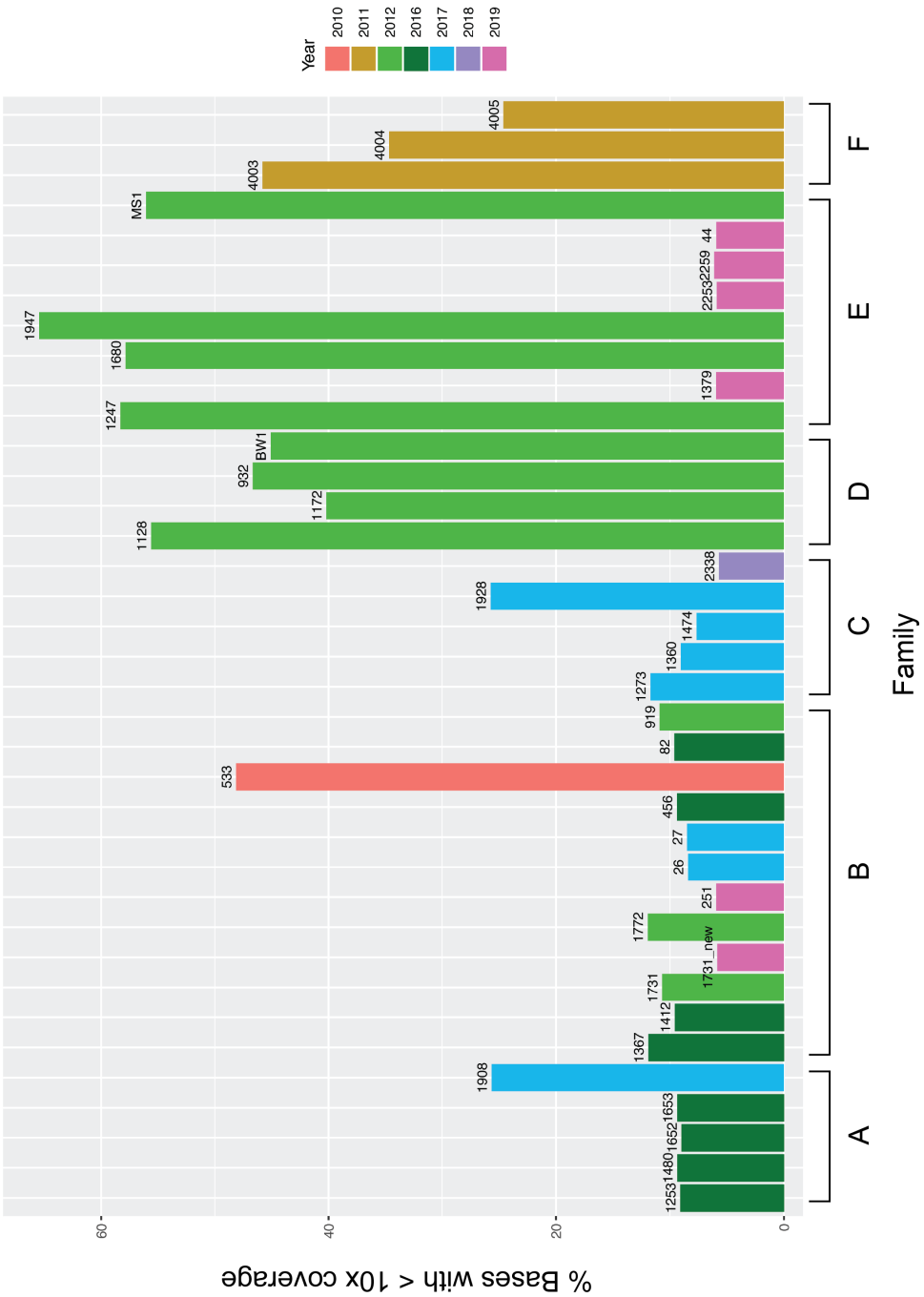

Figure 3.2: Proportion of WES kit target bases with low coverage for all WES samples
Low coverage is defined as less than 10x reads, given as a percentage of the total number of bases within the WES kit target region.

3.3 Variant filtering

Variant filtering can be used in various ways to establish criteria for retention or removal of variants. The following are general types of variant filtering which were used in this project:

- **Genotype**. Retaining variants shared only among certain individuals. For example, variants present only within individuals sharing a phenotype (such as MHS).
- **Functional effect**. Retaining only certain types of variants. For example, variants that are predicted to lead to a protein sequence change.
- **Allele frequency**. Retaining only those variants present at a given population allele frequency (AF). For example, variants that are present in 1% or less of the population.
- **Pathogenicity prediction**. Retaining only variants which meet a certain threshold for prediction of pathogenicity/deleteriousness.
- **Tissue expressivity**. Retaining only variants within genes expressed in a given tissue.
- **Gene**. Retaining variants based on the gene role or the gene's presence in a list of genes of interest.
- **Position**. Retaining only variants present in a given region within the genome. For example, variants which fall within an MH-susceptibility locus. Described further in Section 3.4.2.

Specific details of variant filtering, using the types of filtering described above, are given below. Unless otherwise stated, all variants presented in this work were obtained from filtering in the manner described in the sections below.

3.3.1 Filtering by genotype

An effective method of variant filtering for the discovery of variants associated with a phenotype is through genotype filtering. This involves inferring the expected carriers (the genotype) of a disease-pathogenic variant based on phenotypic information. In the case of a highly penetrant, autosomal-dominant genetic

disorder, the approach to genotype filtering may be clear-cut: all affected individuals would be expected to carry the pathogenic variant, whereas the non-affected individuals would not be expected to carry the pathogenic variant. Therefore, genotype filtering would involve retaining only those variants carried by affected individuals. However, due to the disease aetiology of MH (see Chapter 1.2), some adaptations to genotype filtering were made. Strictly adhering to genotype filtering rules which assumes all MHS individuals harbour the variant and all MHN individuals do not harbour the variant may lead to the exclusion of the MH-pathogenic variant. Family-specific considerations and adaptations to variant analysis were discussed in depth in Chapter 1.6, with a summary of examples of these considerations and the adaptations made presented in Table 1.1.

Initially, VarAFT[123] was used for genotype filtering. However, although VarAFT does allow filtering of variants by genotype, it was determined that it did not allow sufficiently flexible genotype filtering. This flexibility was required for two reasons, one of which was the phenotypic complexity of MH as described above. The second reason was that it was determined that there would be problems caused by variants that were not genotyped for one or more samples within a family. This is because, for a given variant to pass through genotype filtering in VarAFT, each sample must have been genotyped for that variant. A more technical explanation is as follows. In VarAFT, one would set the expected genotype for each disease carrier (MHS) to 'homozygous' or 'heterozygous' (1/1 or 1/0 within the genotype field of a VCF, respectively), and to 'absent' for each MHN sample (0/0 within the VCF genotype field). For a given variant, if any one sample contained a './.' string within the genotype field, meaning that no genotype was inferred (unknown genotype), the variant would automatically be discarded and not proceed through the genotype filter, even if all other genotypes met the programmed conditions. This is problematic, as with increasing numbers of samples within a family VCF, comes an increasing probability that one or more samples is not genotyped for a given variant due to a lack of coverage in that region. This problem is further exacerbated when combining WES samples obtained with different WES kits, as described in Section 3.2.3.

Due to both of these issues, it was necessary to design a system of variant genotype filtering which considered that there were numerous possible combinations of individuals carrying the pathogenic variant within each family, and that there may be individuals present with an unknown genotype. A python

programme, gtFilter, was written to address this.

3.3.1.1 gtFilter for genotype filtering

gtFilter (https://github.com/sormond/gtfilter) is a boolean algebra programme which takes in a VCF and two tab-delimited files: one containing a list of sample names expected to harbour the pathogenic variant ('pos' samples), and one containing a list of sample names not expected to harbour the pathogenic variant ('neg' samples). The output is a VCF containing only those variants meeting those genotype conditions. Additionally, the programme has an optional 'leniency' flag. This flag accepts an integer which represents the number of 'pos' samples that are not required to carry the variant. In the case of MH, this takes into account low specificity rates of the IVCT. However, if there is an individual within the family who may or may not carry an unknown pathogenic variant (for example, an individual who is MHS equivocal or borderline, or an MHS individual who already carries a known variant that may be pathogenic for MH), one would simply not list the sample name on either 'pos' or 'neg' files. This would mean that gtFilter would not take into consideration the genotype of this sample for each variant. The programme automatically passes an unknown genotype of a sample: For example, a variant that has only been called in one and no other samples will pass the genotype filter if the genotype conditions are met for that one sample. The programme has no tolerance for an MHN sample carrying a variant as the IVCT has a very high sensitivity rate and so it is highly unlikely that an MHN individual carries an MH-pathogenic variant. A further explanation of the logic of the programme is presented in Figure 3.3.

3.3.1.2 gtFilter parameters for each family

Table 3.1 gives the gtFilter script 'pos' and 'neg' WES samples for each family, based on the IVCT diagnosis of each individual. Because the IVCT results from individual MS1 (MHS) were borderline (see Appendix A.33 for IVCT values), the MS1 WES sample was not included in either 'pos' or 'neg' lists, meaning the gtFilter programme did not take this genotype into account when filtering variants. IVCT values from individuals 1474 (MHN) and 1928 (MHN) from family C could be considered as borderline, but in this case were still given as disease 'neg' as the IVCT values for all other MHS individuals within the family were much higher. As there is uncertainty about the phenotype of individual 26, WES

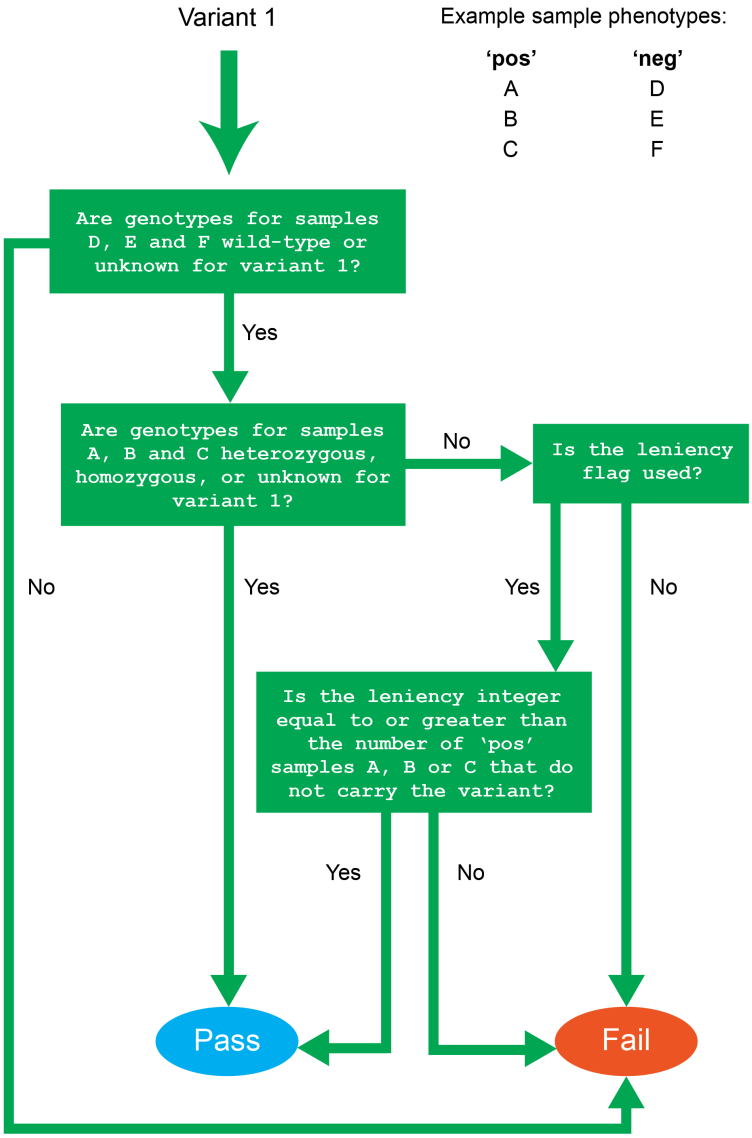

Figure 3.3: Logic of the gtFilter programme for genotype filtering
In this example, three samples (A, B & C) are listed in 'pos', and three others (D, E & F) listed in 'neg'.'Variant 1' is the first variant encountered by the programme. After 'passing' (retaining the variant and adding it to the output VCF) or 'failing' (discarding the variant) variant 1, the programme will then repeat this process on all other variants present within the input VCF.

sample 26 from family B was not included. Additionally, the WES from sample 27 from family B was not included because of individual 27's MHS_h diagnosis, meaning they may not harbour an MH-pathogenic variant.

The 'leniency' argument was used for family B and E due to the larger number of samples and more complex phenotypes present within the families. For both of these, this was set to two, meaning up to two of the 'pos' samples were not required to carry a variant for the variant to pass through the filter.

Family F variants were not genotype-filtered because all WES samples were MHS.

3.3.2 Filtering by functional effect

Filtering by functional effect was carried out in VarAFT. This variant filtering involved selecting only those variants that were categorised as exonic (excluding synonymous) or splicing.

3.3.3 Filtering by allele frequency

Filtering by AF was carried out in VarAFT. The AF threshold was 0.01 (1%) for gnomAD and 1000G AF databases which is conservative as it is above the predicted MH-incidence rate. After this filtering step, variants and all annotations applied by VarAFT were exported from VarAFT. These variant lists will be referred to as *variants of interest*. All subsequent filtering steps were carried out manually on the VarAFT-output variants of interest lists.

3.3.4 Filtering by pathogenicity prediction

VarAFT applies pathogenicity predictions from a large number of pathogenicity predictor tools. No specific threshold was applied for pathogenicity prediction. However, predictions of pathogenicity were taken into account when manually selecting variants of interest. CADD scores (described in Chapter 1.4.4) are given in the final selected variants of interest tables in Chapter 4.

3.3.5 Filtering by tissue expressivity

The Human Protein Atlas[121], a tissue expressivity database, as well as the scientific literature, were queried manually to ascertain whether the gene was

Family	Leniency	Pos	Neg
A	none	1253 1480	1652 1653 1908
B	2	1731 1772 82 1367 533 919 1412 456	251
C	none	1273 1360 2338	1474 1928
D	none	BW1 1172 1128	932
E	2	1947 1247 44 1379 2253	2259 1680

Table 3.1: gtFilter parameters for each family

This table gives the gtFilter python programme parameters for each family. 'Pos' contains the WES samples that were given in the 'pos' file for gtFilter (assumed to carry the unknown pathogenic variant) whereas 'Neg' contains the WES samples that were given in the 'neg' file (assumed to not carry the unknown pathogenic variant). The leniency column gives the leniency parameters for each family, with a value of 'none' meaning the argument was not invoked.

expressed in skeletal muscle. No hard filtering was applied for tissue expressivity because databases are not necessarily complete or accurate, and it is common to find conflicting expression information. Additionally, and especially in the case of families with myopathies, it may be the case that the pathogenic variant is not present within a gene that is expressed in mature skeletal muscle, as the gene may be expressed only early on in myoblast differentiation and turned off after differentiation has occurred.

3.3.6 Filtering by gene

Gene ontology (GO) annotations[116,117], which give information about known and predicted gene function, were applied to the VarAFT-output variant lists using a custom R script (scripts/addGOanno.R). As an example, the GO annotations applied to the *RYR1* gene include 'calcium channel activity', 'calcium ion binding', 'calmodulin binding', 'muscle contraction', 'junctional sarcoplasmic reticulum membrane' and 'release of sequestered calcium ion into cytosol by sarcoplasmic reticulum'. These annotations allowed quick identification of gene roles to ascertain the importance of a variant.

A custom python script was written to query VarAFT-formatted variant lists against gene lists (scripts/geneList.py). A custom curated gene list (available at supp_material/curated_genes.txt), which was expanded upon throughout the project, was queried. This list included genes with an established link or relevance to MH, myopathies, calcium or energy homeostasis in skeletal muscle, ECC, or sarcolemma depolarisation or repolarisation. The list was compiled from both literature searches and personal communications. Additionally, a muscle gene disease list[142] (http://musclegenetable.fr), which was much more comprehensive than the custom curated list, was obtained and queried against (supp_material/NM_disease_genes.txt).

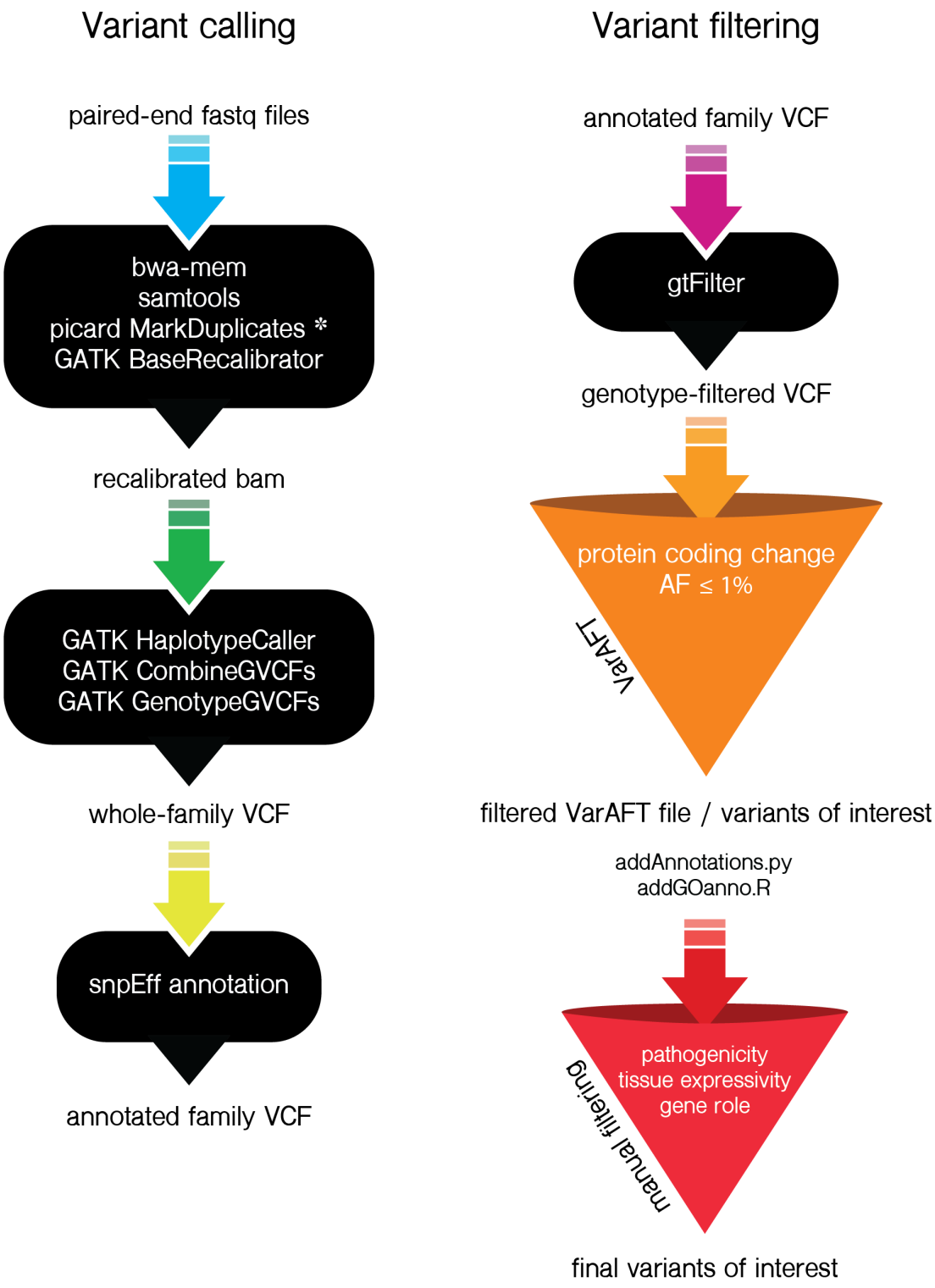

Figure 3.4: Bioinformatic pipeline for variant calling, annotation and filtering
* refers to a step which was not carried out on all WES data (see Chapter 2.9.2).

3.4 Explorations of variant data

Aside from these filtering methods described, variants of interest were explored by identifying variants or genes in common between families and by identifying variants falling within MHS loci.

3.4.1 Shared variants of interest between families

Variants of interest that were shared between families, as well as rare, predicted pathogenic variants within the same gene, were identified. To do this, variants of interest lists for each family were concatenated and sorted based on gene name. The resulting file was manually analysed to identify these variants. Results are reported in Chapter 4.9.

3.4.2 Variants falling within MHS loci

Variants of interest falling within MHS loci (Table 3.2), with the execption of MHS1 and MHS5 which harbour the genes *RYR1* and *CACNA1S*, respectively, were identified manually. Variants within final variants of interest lists presented in Chapter 4 which fall within MHS loci are indicated as such.

Locus	Chr	Position	OMIM
MHS2	17	27,400,000 - 72,900,000	154275
MHS3	7	77,900,000 - 107,800,000	154276
MHS4	3	103,100,000 - 111,600,000	600467
MHS6	5	0 - 48,800,000	601888

Table 3.2: Previously identified MHS loci[46, 47, 48, 49]
'Chr' is chromosome. 'Position' is hg38 genomic coordinates. OMIM is the OMIM accession for the MHS loci.

3.5 Detection of copy-number variation

Identification of copy-number variation (CNV) was carried out using CNVkit. CNVkit is a programme for identification of CNVs using targeted sequencing data. The programme takes in bam files from a 'tumor' sample and a 'normal' sample. The 'tumor' sample can refer to any phenotype not carried by the 'normal' sample. The programme can be run without a 'normal' sample (an MHN individual) if one is not available from a family (it is then said to be compared to a 'flat reference'), but the efficacy of CNV-detection may be reduced and there will be no exclusion of CNVs present within the family that are not related to the disease. CNVkit requires the genome reference and a bed file containing the WES kit target region. It is important that PCR duplicates are marked in the bam files when using CNVkit to assist the programme with detection of true read copy numbers. For WES from all families except family B and E, duplicates had not been marked. Therefore, the bam files of only those WES samples to be used for CNV identification were re-processed with Picard's MarkDuplicates tool as described in Chapter 2.9.8. It is also important that WES samples that were compared were also sequenced with the same sequencing method and with the same WES kit. Where possible, WES samples with the highest average coverage within the family, and from individuals that were not considered borderline MHS or MHN, were used.

3.5.1 WES samples used for CNV detection

Table 3.3 gives the WES samples used for CNV analysis within each family. For family B, two WES samples from MHS individuals were screened for CNVs: individuals 1731 and 1412. As 1731 was MGI-sequenced, the WES sample from MHN individual 251, also MGI-sequenced, was used as the 'normal' sample. There was no suitable MHN WES sample to use as the 'normal' sample for sample 1412, and so a flat reference was used. For family F, the WES sample of individual 4004 (MHS) was used against a flat reference as no MHN WES sample was available from this family.

Family	MHS	MHN
A	1480	1652
B1	1731	251
B2	1412	flat reference
C	1273	1474
D	1172	932
E	2253	2259
F	4004	flat reference

Table 3.3: WES samples used for CNV detection from each family
'B1' is family B MGI-sequenced samples, 'B2' is family B Illumina-sequenced samples.

3.5.2 Analysis of predicted copy-number variation

If CNVkit suspected the presence of a single copy deletion, evidence of SNP heterozygosity over the suspected deleted area upon visual inspection of the alignment often ruled this out. Additionally, the presence of an alternative haplotype within the suspected deleted region would often explain the detection of a deletion, as the reads have likely mapped to the alternative chromosome.

To visualise coverage of the genomic regions containing a suspected CNV, plots were made in R using the Sushi package[138] with mosdepth per-base coverage metrics.

3.6 Family F special case

Because family F had no MHN WES samples, an additional filtering step was applied to reduce the final number of variants to be analysed. snpSift[106] (version 4.3t, part of snpEff) was used to filter the snpEff annotated family VCF to retain only those variants predicted to cause a protein coding change (see Chapter 2.9.7 for method). This was done to significantly reduce the number of variants needed to be analysed within VarAFT, which cannot handle large variant numbers. Filtered VCFs were then annotated and analysed using VarAFT, with an additional step whereby variants were filtered so that only variants heterozygous or homozygous in all samples were retained.

Chapter 4

Variants of interest

This chapter presents variants that were identified in each family which may be related to their malignant hyperthermia (MH) and/or myopathy phenotype. Although segregation analysis was performed on select variants of interest, testing segregation of all variants was beyond the scope of this project. Rare variants that are shared among families, or genes containing rare variants shared among families, are also given here, even if they do not have an obvious link to the current biomolecular understanding of MH.

4.1 Methods

For detailed methods and method development, see Chapter 3. Unless otherwise stated, variant lists given here are final selected variants of interest. This means that the variants segregate with the disease according to the WES data (the details of expected segregation differs amongst families, as described in Chapter 1.6 & section 3.3.1), that they are present at a gnomAD allele frequency (AF) of ≤ 0.01, and that they are predicted to alter the protein sequence of a gene's encoded protein. Additionally, they have been selected for the final list of variants of interest because they have some relevance to skeletal muscle, calcium homeostasis, energy metabolism, and/or muscle disease. Some variants with relevance to skeletal muscle do not have a clear relevance to the understanding of MH pathophysiology, but are presented because no other suitable candidate variants were identified within the family.

For primer sequences used for testing of segregation of the variants, see Appendix A.1.

4.2 Number of identified variants

Table 4.1 gives variant numbers for each family at each major stage of variant filtering.

Family	**VCF**	**gtFiltered**	**<0.01 AF & protein changing**	**VOI**
A	430897	90601	156	6
B	3052842	2201899	145	7
C	1075991	563575	142	4
D	2264434	872848	246	3
E	2888835	1341484	221	4
F	4413113	N.A.	74	2

Table 4.1: Variant numbers at each major stage of the variant filtering pipeline 'VCF' is the number of variants in the whole family VCF. 'gtFiltered' is the number of variants after applying genotype filtering using the gtFilter python script. '<0.01 AF & protein changing' is variant numbers after VarAFT filtering to remove variants that have an AF of greater than 0.01 and variants that are not predicted to alter the gene's protein product. 'VOI' (variants of interest) are final VOI that were manually selected from the VarAFT filtered variant list and that are given in this chapter.

4.3 Family A

Six final variants of interest identified are listed in Table 4.2.

Chr	Gene	Protein	Role	Variant	AA change	dbSNP(147)	gnomAD AF	CADD
4	*FAT1*	FAT1	cell-cell adhesion, morphogenesis of neuromuscular junction [143, 144]	NM_005245:exon8:c.4434C>G	I1478M	rs376344008	1.63E-05	23.3
4	*PPARGC1A*	PGC1α	regulation of genes involved in skeletal muscle energy metabolism[145]	NM_013261:exon3:c.412G>T	A138S	.	.	23.9
4	*TLR3*	TLR3	myocyte inflammatory response[146, 147]	NM_003265:exon4:c.958G>A	G320R	rs554658931	1.63E-05	25.4
5	*PRKAA1*	AMPK α1 subunit	regulation of metabolism in response to cellular energy states[148]	NM_006251:exon7:c.1286A>G	K429R	rs146544425	0.0005	18.6
6	*SLC22A3*	Oct3	organic cation transport[149]	NM_021977:exon1:c.131C>T	T44M	rs8187715	1.74E-05	22.9
9	*VPS13A*	VPS13A	endosomal protein sorting, mitochondrial size maintenance[150]	NM_001018037:exon35:c.4036T>G	S1346A	.	.	7.946

Table 4.2: Family A final variants of interest
'AA' is amino acid. 'AF' is allele frequency. 'CADD' is the reported CADD score for the variant. Variant descriptions and amino acid change are the ones first reported by ANNOVAR and are based on the Refseq gene accession given, which is generally, but not necessarily, the canonical isoform.

4.3.1 *FAT1* I1478M SNV

A SNV in *FAT1*, predicted to lead to an isoleucine to methionine substitution at amino acid position 1478 (Refseq accession: NM_005245), was identified in family A. The variant is predicted to be pathogenic (CADD score: 23.3).

The encoded protein of *FAT1* (OMIM: 600976) is a member of the cadherin family. The cadherins are a group of proteins which are involved with cell-to-cell adhesion. The protein is known to contribute to morphogenesis of the neuromuscular junction[143] and was shown in a mouse knock-out study to be critical for the correct formation of muscle[144]. This is of interest as incorrect functioning of the neuromuscular junction could conceivably contribute to MH-susceptibility. Some variants in the gene, thought to alter expression of the gene product, have been linked to facioscapulohumeral muscular dystrophy[151], a genetic disorder characterised by progressive skeletal muscle weakness[152], demonstrating that *FAT1* may be important for normal skeletal muscle function.

4.3.2 *SLC22A3* T44M SNV

A *SLC22A3* SNV identified in family A is predicted to lead to an alanine to serine substitution at amino acid position 138 (Refseq accession: NM_021977). The variant is predicted to be pathogenic (CADD score: 22.9).

SLC22A3 encodes Oct3, an extraneuronal organic cationic transporter protein. Variants in the gene have not been previously linked to disease, and little is known about the protein's function, except that it mediates transport of a variety of non-charged and cationic organic molecules[149]. However, the variant is present within the final variants of interest list as Oct3 is highly expressed in skeletal muscle[153] ,[121].

4.3.3 *VPS13A* S1346A SNV

A SNV in *VPS13A*, predicted to lead to a serine to alanine substitution at amino acid position 1346 (Refseq accession: NM_001018037), was identified in family A. The variant appears to be novel as it was not found on any database. It is not predicted to be pathogenic (CADD score: 7.946).

VPS13A encodes a protein involved in lyososomal and endosomal cycling of

proteins which has been implicated in mitochondrial size maintenance and clearance[150]. It is expressed highly in a number of tissues including skeletal muscle. Variants in *VPS13A* have been linked to choreoacanthocytosis, a neurological disorder with symptoms including myopathy and involuntary muscle movements[154]. This variant is of interest because of the mitochondria's key role in energy homeostasis: It is plausible that dysfunction of the mitochondria leading to energy dysregulation, such as in an 3,4-Methylenedioxymethamphetamine (MDMA)-induced hyperthermic episode, could have a role in MH.

4.3.4 *PPARGC1A* A138S SNV

A SNV within *PPARGC1A*, predicted to lead to an alanine to serine substitution at amino acid position 138 (Refseq accession: NM_013261), was identified in family A. The variant appears to be novel as it could not be found on any database. It is predicted to be pathogenic (CADD score: 23.9).

PPARGC1A encodes a well-studied muscle transcriptional cofactor, PGC1α. The protein is expressed highly in skeletal muscle, where it controls the expression of genes involved in energy metabolism[145]. PGC1α is understood to be involved in muscle fibre-type switching (transitioning between fast and slow-twitch muscle fibre) and regulation of mitochondrial biogenesis[145].

Amino acid position 138 of the protein resides in the activation domain of PGC1α, responsible for interactions with other co-factors and docking of nuclear receptors[155]. An amino acid substitution in this position could potentially alter the strength of interactions with these binding partners, which could effect expression of the genes under the control of PGC1α.

The variant was selected for segregation analysis due to the gene's key role in muscle metabolism through control of gene expression, and because the variant is novel.

4.3.4.1 Segregation analysis

Sanger sequencing was carried out on the two remaining family A members for which gDNA and IVCT results were available. Although sequencing was repeated a number of times due to the presence of contaminating sequence (Figure 4.1), the results indicated that MHN 1576, child of 1479, carried the variant. The variant was excluded from future analyses based on this, however, sequencing should be

repeated using another set of sequencing primers to confirm that individual 1576 does carry the *PPARGC1A* variant.

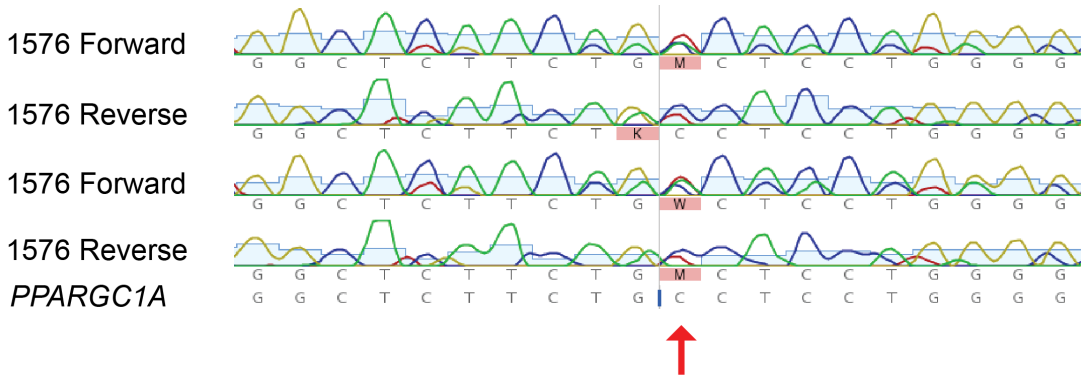

Figure 4.1: Alignment of Sanger sequencing results for individual 1576 against the *PPARGC1A* nucleotide sequence

The *PPARGC1A* A138S SNV position is indicated by the red arrow. Although there appears to be some sequence contamination, the results suggest that this individual is likely to be heterozygous, with peaks for both C and A nucleotides present in all four Sanger reads.

4.3.5 *TLR3* G320R SNV

A *TLR3* variant identified in family A, predicted to cause a glycine to arginine substitution at amino acid position 320 (Refseq accession: NM_003265), is predicted to be pathogenic (CADD score: 25.4). The amino acid at this position is highly conserved (Figure 4.3).

Toll-like receptor 3 (TLR3), encoded by the gene *TLR3* (NG_007278.1:5001-20944), is a transmembrane receptor protein with a well-studied role in recognition of pathogen-associated molecular patterns and subsequent activation of the innate immune system. This activation results in the induction of type I interferons and inflammatory cytokines[156], leading to both an inflammatory and immune response. TLR3 primarily localises to endosomal and endoplasmic reticulum membranes.

Aside from a role in innate immune system activation, TLR3 has been more recently implicated as a mediator of the myocyte inflammatory response, the maturation of cardiac myocytes[157], and the regulation of cardiac potassium channels responsible for action potential regeneration. Although expression of *TLR3* has not been observed in skeletal muscle myocytes, it is very highly expressed in myoblasts and in the muscle tissue of patients suffering from inflammatory and

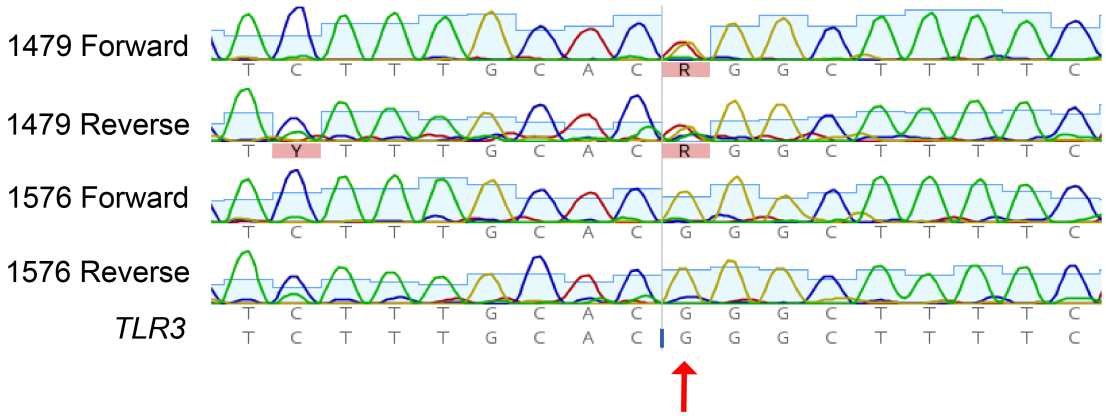

Figure 4.2: Alignment of Sanger sequencing results for individuals 1479 and 1576 against the *TLR3* nucleotide sequence

The *TLR3* G320R SNV position is indicated by the red arrow. The results indicate that individual 1479 is heterozygous for the variant, whilst individual 1576 does not carry the variant.

HIV-myopathies[146, 147].

In the cardiac tissue of mice, knock-out of *TLR3* has been demonstrated to affect the generation of an action potential by lengthening the time between depolarisation and repolarisation of the membrane[158]. This effect was shown to be independent of the inflammatory and immune signalling functions of TLR3, but dependent on direct regulation by TLR3 on the proteosomal degradation of subunits of the cardiac transient outward potassium channel, Kv4.2/4.3.

The variant was selected for segregation analysis due to the gene's encoded protein's role in myocyte inflammatory response, myoblast expression, and because the protein has been implicated in the polarisation of the membrane in cardiac muscle.

4.3.5.1 Segregation analysis

Individuals 1479 (MHS) and 1576 (MHN) were Sanger sequenced for the variant which revealed segregation with MH susceptibility (Figure 4.2). Further study of the variant has not yet been undertaken. The next course of action would be to establish an HRM assay to screen a population of MHN individuals for the variant.

Figure 4.3: Conservation of TLR3 glycine amino acid at position 320 amongst species

Other species are mouse, rat, bovine, antelope, guinea pig and dog.

4.3.6 *PRKAA1* K429R SNV

A *PRKAA1* SNV, predicted to lead to a lysine to arginine substitution at amino acid position 429 (Refseq accession: NM_006251), was identified in family A. The variant falls within the MHS6 locus. It is predicted to be pathogenic (CADD score: 18.6).

PRKAA1 encodes the α1 catalytic subunit of AMP-activated protein kinase (AMPK), an energy-sensing protein kinase which regulates metabolism in response to cellular energy states and intracellular Ca^{2+} concentrations[148]. AMPK is composed of three subunits: α, β and γ. These subunits are each encoded by multiple isoforms, allowing tissue-specific variation of the properties of AMPK[159]. In skeletal muscle, α2 is the predominant α isoform, with moderate expression of α1[121 159]. However, α2 is thought to have a nuclear role involving gene regulation, whilst α1 remains within the cytoplasm[160].

AMPK has an important role in myocyte energy homeostasis, and because energy states contribute to the regulation of RyR1, the variant was chosen for segregation analysis.

4.3.6.1 Segregation analysis

The WES data indicated that the variant was heterozygous in individuals 1480 and 1253 and absent from MHN individuals 1652 and 1653. WES coverage at the variant site was low (depth of five) for MHN sample 1908, however the reads indicated that the variant was not present in this individual (Appendix A.2). IVCT results and DNA were available from only two other individuals within family A. Sanger sequencing revealed segregation of the variant with MHS (Appendix A.3).

Next, an HRM assay was developed to screen the MHN population for the variant. An initial optimisation assay using sequenced samples 1576 and 1479 gave a reasonable heteroduplex melting peak (Appendix A.4). The assay was carried out on DNA from the five other family A individuals to confirm the genotypes suspected through WES and to assess viability of the assay on a larger number of samples (Appendix A.5). The melting curves suggested that MHN 1908, presumed wild-type from the WES data, was a carrier of the variant. Suspected wild-type 1653 gave an unusual melting curve with a wide and small secondary peak which suggested the individual to be a carrier of the variant. Alternatively, gDNA degradation may have caused the secondary peak. Individuals 1908, 1653 and 1480 were Sanger sequenced to confirm their genotypes. The sequencing revealed MHN 1908 to carry the variant, confirmed MHS 1480 as a carrier and confirmed individual 1653 as wild-type for the variant (Figure 4.4). The finding that 1908 was a carrier of the variant demonstrated how low sequence coverage can conceal heterozygosity of a DNA sample, potentially misguiding future research approaches. It provides a good example of allelic dropout in WES, and underscores the importance of undertaking confirmatory Sanger sequencing as early as possible if there is any ambiguity in WES data.

Overall, the finding that MHN individual 1908 carried the variant disproved the possibility that the variant is pathogenic for MH, and the variant was not pursued further.

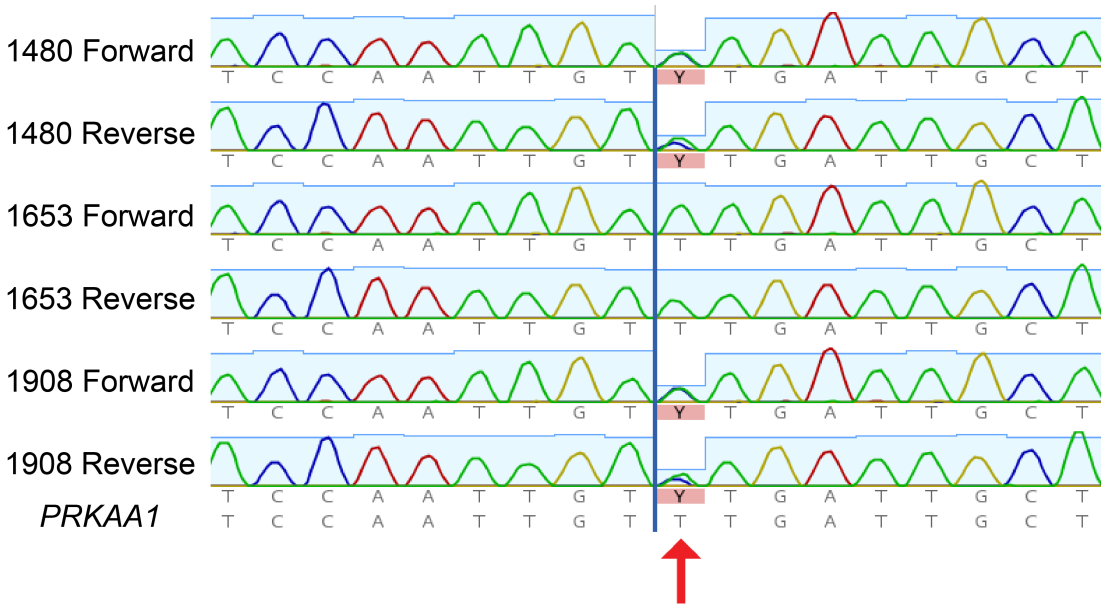

Figure 4.4: Alignment of Sanger sequencing results for individuals 1480, 1653 and 1908 against the *PRKAA1* nucleotide sequence
The *PRKAA1* K429R SNV position is indicated by the red arrow. The results indicate that individuals 1480 and 1908 are heterozygous for the *PRKAA1* variant whereas individual 1653 does not carry variant.

4.4 Family B

A list of variants of interest can be found in Table 4.3.

Chr	Gene	Protein	Role	Variant	AA change	dbSNP(147)	gnomAD AF	CADD
1	*KLHL21*	KLHL21	formation of Cullin E3 ubuiqitin ligase complex[161]	NM_001324309:exon1:c.590C>T	A197V	rs369534954	0.0031	15.6
13	*LMO7*	LMO-7	transcriptional regulator of muscle-specific genes[162]	NM_001366634:exon15:c.2344C>G	Q782E	rs564271731	1.26E-05	13.51
14	*AHNAK2*	AHNAK2	possible cytoskeletal role[163]	NM_001350929:exon7:c.5555C>T	P1852L	rs79599750	0.0038	19.75
15	*PPIP5K1*	PPIP5K1	inositol pyrophosphate cellular signalling[164]	NM_001354395:exon27:c.3242G>A	R1081H	rs143339052	0.0001	26.1
17	*CDRT1*	CMT1A duplicated region transcript 1 protein	possible ubquitination role[165]	NM_001282540:exon5:c.1118C>T	T373M	rs200713812	0.0002	24.8
17	*CDRT1*	see above	see above	NM_001282540:exon1:c.245G>A	G82E	.	.	24.5
17	*TRPV2*	Trpv2	Ca^{2+} signalling in response to various effectors[166]	NM_016113:exon4:c.586C>T	R196C	.	1.65E-05	24.3

Table 4.3: Family B final variants of interest
'AA' is amino acid. 'AF' is allele frequency. 'CADD' is the reported CADD score for the variant. Variant descriptions and amino acid change are the ones first reported by ANNOVAR and are based on the Refseq gene accession given, which is generally, but not necessarily, the canonical isoform.

4.4.1 *KLHL21* A197V SNV

A SNV within *KLHL21*, predicted to lead to an alanine to valine subsitution at amino acid position 197 (Refseq accession: NM_001324309), was identified in family B. The variant is predicted to be pathogenic (CADD score: 15.6).

KLHL21, encoding KLHL21, belongs to a large protein family characterised by the presence of kelch-repeat domains which are associated with protein-protein interactions. Expression of KLHL21 is ubiquitous but it may be most highly expressed in skeletal muscle[121]. A number of proteins belonging to the kelch-repeat domain family have important roles in skeletal muscle development and maintenance, and variants in some of these genes, including *KLHL1*, *KLHL9*, *KLHL16*, *KLHL40* and *KLHL41*, have been implicated in a number of neuromuscular disorders[167]. A precise role for KLHL21 in skeletal muscle has not been determined, however it is known to form a part of Cullin E3 ubquitin ligase complexes, which are involved in protein degradation and turnover[161]. The variant is of interest due to the fact that variation in this family of genes are pathogenic for neuromuscular disorders. It is plausible that erroneous protein degradation could lead to incorrect functional expression of proteins involved in calcium homeostasis or excitation-contraction coupling (ECC).

4.4.2 *LMO7* Q782E SNV

A *LMO7* SNV, predicted to lead to a glutamine to glutamic acid subsitution at amino acid position 782 (Refseq accession: NM_001366634), was identified in family B. The variant has mixed pathogenicity predictions (CADD score: 13.51).

LMO7 encodes a protein which transcriptionally regulates a number of muscle genes including emerin, a nuclear envelope protein critical for anchoring of the nuclear membrane to the cytoskeleton[162]. Variants of the gene encoding emerin are associated with Emery-Dreifuss muscular dystrophy[168]. No variants of *LMO7* have been linked to disease, but a phenotype similar to Emery-Dreifuss muscular dystrophy was observed in mice with an *LMO7* knock-down[169], although these findings could not be replicated in another similar experiment[170]. The variant is of interest due to the gene's association with a disorder of skeletal muscle.

4.4.3 *AHNAK2* P1852L SNV

An *AHNAK2* SNV, predicted to lead to a proline to leucine acid substitution at amino acid position 1852 (Refseq accession: NM_001350929), was identified in family B. The variant is predicted to be pathogenic (CADD score: 19.75).

AHNAK2 encodes AHNAK2, a member of the AHNAK family. The AHNAK family also includes the protein AHNAK. AHNAK and AHNAK2 are closely related, with both sharing significant sequence identity, size (600 kDa) and structural features[171]. AHNAK2 is less studied than AHNAK, but both are believed to share cellular roles, as mouse knock-out studies have suggested AHNAK2 compensates for the absence of AHNAK[171]. In skeletal muscle, AHNAK and AHNAK2 localise to costameres, muscle cell structures which connect sarcomeres to the sarcolemma[163]. In cardiac tissue, AHNAK is a modulator of the voltage-dependent cardiac calcium channel $Ca_v1.2$[172]. Both AHNAK and AHNAK2 have been shown to interact with JP-45[173]. The variant is of interest because this gene is highly expressed in skeletal muscle and it may be an interacting partner of JP-45, an important ECC protein.

4.4.4 Compound *CDRT1* variants

Two rare variants were identified in *CDRT1*, predicted to lead to threonine to methionine, and glycine to glutamic acid substitutions at amino acid positions 373 and 82 (T373M & G82E), respectively (Refseq accession: NM_001282540). Both variants are predicted to be pathogenic (CADD scores: 24.8 and 24.5). The carrier status of the variants, inferred from the WES data, suggests the variants segregate together, meaning carriers are cis compound heterozygotes. All WES samples appear to carry both variants with the exception of the WES sample from individuals 251 (MHN), 26 (MHN) and 919 (MHS), who are all wild-type for both variants. WES sample 82 was called as wild-type for T373M and so is non-concordant with the inferred haplotype (17 reads containing wild-type sequence, 2 reads with variant sequence), however is likely a carrier as this individual links both branches of the family and appears to carry the *CDRT1* G82E variant. This implies that allele dropout may have prevented identification of this variant in individual 82.

Duplications of chromosomal segment 17p11.2p12, which harbours *CDRT1*

and the gene *PMP22*, are causative of the neuropathy Charcot-Marie Tooth syndrome (CMT)[174]. It is worth noting however that the duplication of *PMP22* seems to be the causative agent of CMT, with point variants in *PMP22* associated with CMT, although this doesn't rule out a pathogenic role for *CDRT1*. Sufferers of CMT may be at risk of MH[175]. Aside from its relationship to this disease, little is known about the function of *CDRT1*, although analysis of its predicted protein structure implicates it in protein ubiquitination (as predicted by Uniprot)[165]. Although no recent literature focused on the gene exists, the original literature suggests high and specific skeletal muscle RNA expression[176]. These variants are of interest due to the gene's potential role in a neuropathy which may put sufferers at risk of MH, and because two predicted pathogenic variants were identified within the same gene.

4.4.5 *PPIP5K1* R1081H SNV

A *PPIP5K1* SNV, predicted to lead to an arginine to histidine substitution at amino acid position 1081 (Refseq accession: NM_001354395), was identified in family B. The variant is predicted to be pathogenic (CADD score: 26.1).

PPIP5K1 encodes diphosphoinositol pentakisphosphate kinase 1 (PPIP5K1), an enzyme involved in inositol pyrophosphate cellular signalling. Specifically, PPIP5K1 catalyses the conversion of inositol hexakisphosphate to diphosphoinositol pentakisphosphate, and diphosphoinositol pentakisphosphate to bis-diphosphoinositol tetrakisphosphate[164]. Inositol pyrophosphates are highly phosphorylated signalling molecules which are involved in the regulation of a variety of cellular processes including glucose homeostasis, stress response, protein translation and phosphate homeostasis[177,178]. Inositol pyrophosphate signalling has recently been implicated in alterations of calcium transients[179]. *PPIP5K1* is expressed in skeletal muscle[180].

The variant was chosen for segregation analysis due to the gene's potential involvement in cellular signalling and metabolic control of myocytes, and because the variant is predicted to be pathogenic.

4.4.5.1 Segregation analysis

Sanger sequencing was carried out on individuals 1412 and 1731 which confirmed the genotypes suspected by WES (Appendix A.6). An initial HRM optimisation was carried out, which gave expected melting curves for wild-type 1731 and heterozygous 1412 (Appendix A.7). This HRM assay was then carried out on both MHN and MHS individuals from family B, which revealed that the variant did not segregate with MHS (Figure 4.5). The variant was excluded from further analysis.

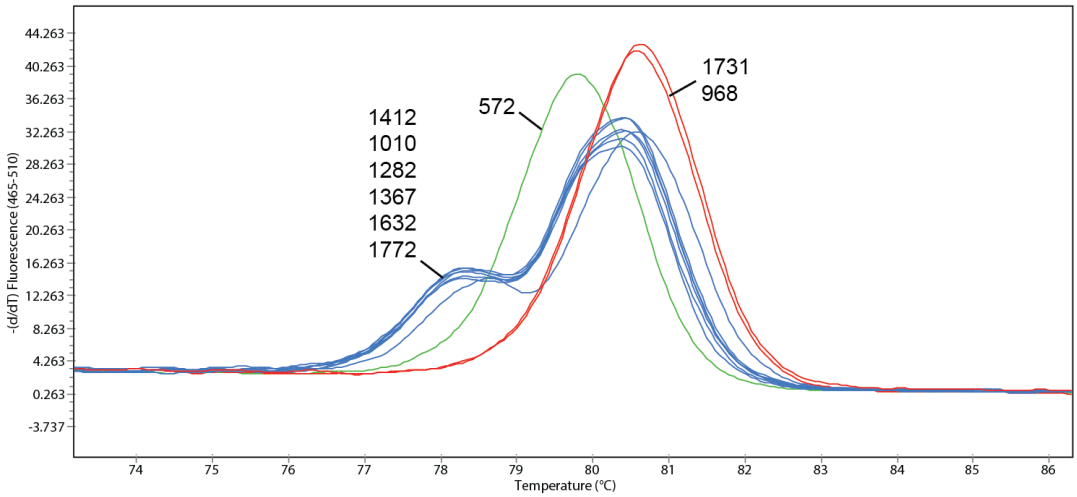

Figure 4.5: HRM melting curve for *PPIP5K1* R1081H variant
1412 is confirmed heterozygous for the variant and 1731 is confirmed wild-type. The results indicate that individuals 1010 (MHS_c), 1282 (MHN), 1367 (MHS), 1632 (MHN) and 1772 (MHS) are heterozygous for the variant, whereas individual 968 (MHS) is wild-type for the variant. The genotype for individual 572 (MHS) is inconclusive as the single homoduplex peak is shifted from the control wild-type peak.

4.4.6 *TRPV2* R196C SNV

A *TRPV2* variant, predicted to cause an arginine to cysteine substitution at amino acid position 196 (Refseq accession: NM_016113), was identified in family B.

TRPV2 encodes Trpv2, which belongs to the TRP ion channel protein family, and more specifically to the transient receptor vanilloid (TRPV) family. *TRPV2* is expressed ubiquitously with some expression in skeletal muscle[121]. The primary function of TRPs is the mediation of sensory signalling by rapid ion influx or

efflux in response to sensing of a specific signal. TRPVs are selective for Ca^{2+} and form homotetrameric channels, with evidence for coassembly of different TRPV subunits to form heterotetrameric channels[166]. *TRPV1*, which encodes Trpv1, is closely related to Trpv2, sharing over 50% sequence identity[40]. Recently, it has been proposed that Trpv1 may contribute to MH susceptibility[181], with evidence suggesting that Trpv1 is activated by volatile anaesthetics in human skeletal muscle. The researchers found that in a mouse model SR calcium release upon exposure to the volatile anaesthetic isoflurane was increased significantly by the presence of Trpv1 channels harbouring variants found in MH families[181]. This provides evidence for non RyR1-specific calcium release from the SR upon exposure to an MH-triggering agent, and potentially implicates TRPV channels.

The variant was chosen for segregation analysis due to the gene's potential role in myocyte Ca^{2+} homeostasis.

4.4.6.1 Segregation analysis

Sanger sequencing was carried out on individual 1367 (MHS) which confirmed the presence of the heterozygous variant (Appendix A.8). An initial HRM optimisation assay with 1367 gDNA and an MHN sample gave characteristic wild-type and heterozygous melting curves (Appendix A.9) and so this HRM assay was carried forward for genotyping of the wider family for the *TRPV2* R196C variant. This revealed discordance of the variant with MH susceptibility, with an MHN individual (740) found to harbour the variant, and MHS individuals determined to be wild-type for the variant (Figures 4.6 & 4.7).

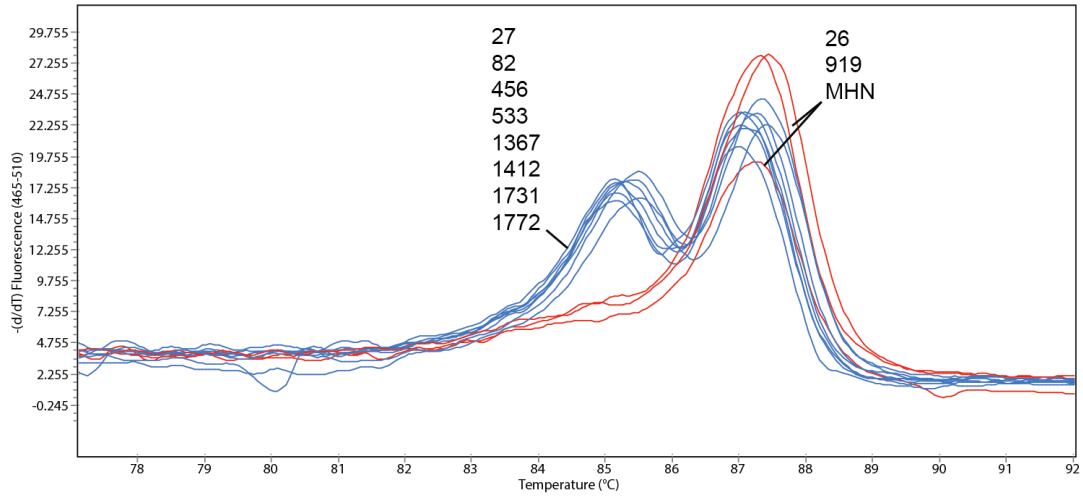

Figure 4.6: HRM melting curve for *TRPV2* R196C variant
1367 (MHS) is confirmed heterozygous for the variant and 26 is confirmed wild-type (Appendix A.8). The results indicate that individuals 27 (MHS_h), 82 (MHS), 456 (MHS), 533 (MHS), 1412 (MHS), 1731 (MHS) and 1772 (MHS) are heterozygous for the variant, whereas individuals 26 (uncertain phenotype), 919 (MHS) and MHN are wild-type for the variant.

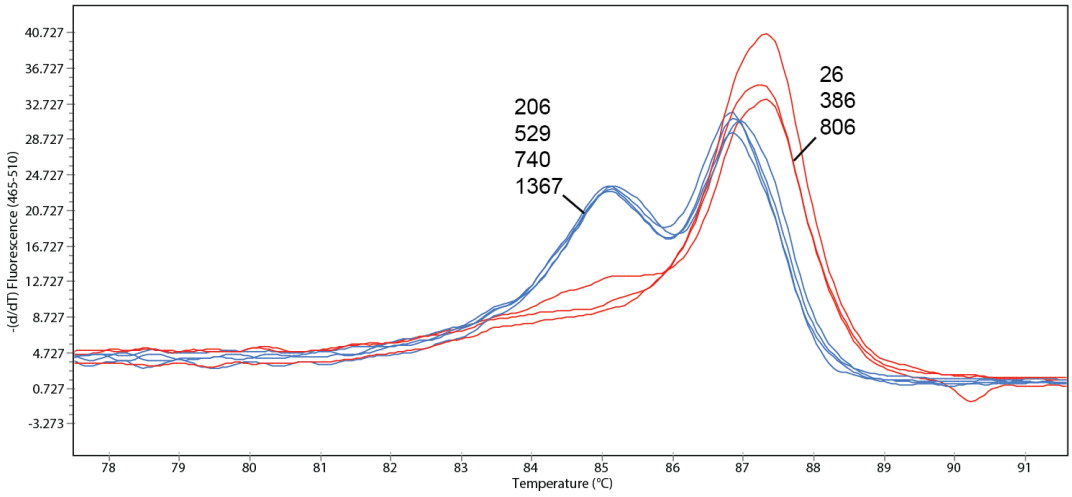

Figure 4.7: HRM melting curve for *TRPV2* R196C variant
1367 is confirmed heterozygous for the variant and 26 is confirmed wild-type. The results indicate that individuals 206 (MHS), 529 (MHS) and 740 (MHN) are heterozygous for the variant, whereas individuals 386 (MHN) and 806 (unknown phenotype) are wild-type for the variant.

4.5 Family C

A list of variants of interest can be found in Table 4.4.

Chr	Gene	Protein	Role	Variant	AA change	dbSNP(147)	gnomAD AF	CADD
16	*SLC6A2*	NET	norepinephrine transport at synaptic clefts[182]	NM_001172502:exon10:c.1283T>C	V428A	.	.	24
16	*PDPR*	PDPR	regulation of pyruvate dehydrogenase complex[183]	NM_001322118:exon4:c.265G>A	A89T	rs202112115	0.0008	13.1
16	*ZFHX3*	ATBF1	transcriptional regulator of muscle-differentiation factors[184]	NM_006885:exon2:c.2330_2347del	777_783del	rs762509156	0.0003	.
18	*LRRC75B*	leucine-rich repeat-containing protein 75B	possible role in myocyte differentiation[185]	NM_207644:exon4:c.821C>T	P274L	rs202090927	0.0007	29.5

Table 4.4: Family C final variants of interest
'AA' is amino acid. 'AF' is allele frequency. 'CADD' is the reported CADD score for the variant. Variant descriptions and amino acid change are the ones first reported by ANNOVAR and are based on the Refseq gene accession given, which is generally, but not necessarily, the canonical isoform.

4.5.1 *SLC6A2* V428A SNV

A variant in *SLC6A2* identified in family C, predicted to lead to a valine to alanine substitution at amino acid position 428 (Refseq accession: NM_001172502), is believed to be novel, as the variant was not present on any database. It is predicted to be pathogenic (CADD score: 24).

SLC6A2 encodes the norepinephrine transporter (NET). NET is responsible for the re-uptake of extracellular norepinephrine (NE) at synaptic clefts[182]. NE elicits the 'fight or flight' response, triggering a range of physiological effects which includes the promotion of glucose uptake in skeletal muscle[186]. In MDMA-induced hyperthermia, it is thought that an elevation of NE triggered by MDMA goes on to indirectly elicit an over-activation of a mitochondrial inner membrane protein, UCP3[74]. UCP3, a protein expressed highly and specifically in skeletal muscle[121] mediates the passive leakage of protons across the inner mitochondrial membrane, diffusing the proton gradient. ATP is expended and heat is produced by this futile cycle, causing the symptoms of MDMA-induced hyperthermia. For further discussion on MDMA-induced hyperthermia and a link to another variant identified in family D, see section 4.6.3. The *SLC65A2* variant is of interest due to the fact that it is novel and because the gene is involved in a physiological pathway which is implicated in triggering a physiological response similar to a MH episode.

4.5.2 *PDPR* A89T SNV

A *PDPR* variant, predicted to lead to an alanine to threonine substitution at amino acid position 89 (Refseq accession: NM_001322118), was identified in family C. The variant is not predicted to be pathogenic (CADD score: 13.1).

This variant is of interest due to the gene's role in energy metabolism. *PDPR* encodes the pyruvate dehydrogenase regulatory subunit, or PDPR, which makes up part of the pyruvate dehydrogenase complex (PDC). The PDC carries out the key reaction linking glycolysis and the TCA cycle - that is, the oxidative carboxylation of pyruvate forming acetyl-CoA. PDPR decreases the sensitivity of PDC to stimulatory Mg^{2+} [183]. *PDPR* is ubiquitously expressed throughout the body[121]. Genetic variation in this gene has not previously been shown to be pathogenic for disease.

4.5.3 *ZFHX3* 18-nucleotide deletion

A *ZFHX3* variant identified in family C is an 18 nucleotide in-frame deletion, predicted to lead to a deletion of 5 amino acids (Val-Ala-Ala-Ala-Ala) from amino acid position 777 (Refseq accession: NM_006885).

ZFHX3 encodes ATBF1, a transcriptional regulator of myogenic and neuronal differentiation[184]. Decreasing expression of *ZFHX3* correlates with differentiation of myoblasts into terminal muscle cells[184]. ATBF1 may be involved in regulation of calcium homeostasis through control of gene expression. Using a mammary epithelial cell line, ATBF1 was shown to regulate the expression of the calcium channel *TRPV6* and its knock-down led to a decrease in store-operated calcium entry, whilst its over-expression had an opposing effect[187]. No *ZFHX3* variant has been associated with rare genetic disease. However, the gene has been implicated in atrial fibrillation (AF), a disturbance of cardiac rhythm. Polymorphisms of the gene, including intronic variants, have been associated with an increased risk of AF[188]. In an atrial myocyte cell line, knock-down of *ZFHX3* caused a dysregulation of calcium homeostasis, with an increase in SR calcium store content, increased expression of calcium homeostasis proteins including RYR2 and SERCA2, and increased activity of channels involved in action potential generation[189].

The deletion variant falls within an alanine-repeat region of the protein. The predicted alteration of amino acid sequence from wild-type is as follows, with the deleted sequence shown in bold:

wild-type sequence: QVFSHTAGAAAAA**VAAAA**AAANISSS
variant sequence: QVFSHTAGAAAAAAAANISSS

This alanine-repeat region is highly conserved amongst species (Figure 4.8). Other deletion variants in this alanine-rich region that are predicted to lead to the same 5 amino acid deletion, or a similar non-frameshift deletion in the same area, have been identified previously. However, all of these variants have a gnomAD AF of $\leqslant 0.0005$. Due to the presence of central core disease in family C, this variant is interesting as it is conceivable that the dysfunction of an important transcriptional regulator of myogenesis could lead to incorrect expression and/or assembly of the ECC molecular apparatus.

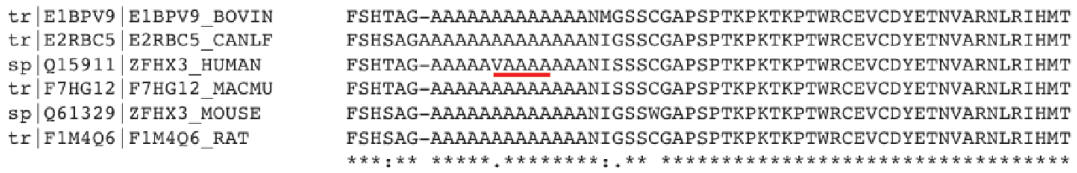

Figure 4.8: Conservation of ZFHX3 amino acid sequence amongst species
The red underlined sequence indicates the deleted amino acid sequence. Other species are bovine, dog, macaque, mouse and rat.

4.5.4 *LRRC75B* P274L SNV

An *LRRC75B* variant was identified in family C. It is predicted to lead to a proline to leucine substitution at amino acid position 274 (Refseq accession: NM_207644). The variant is predicted to be pathogenic by multiple pathogenicity predictors (CADD score: 29.5) and is present at a low AF (gnomAD AF: 0.0007).

LRRC75B encodes a leucine-rich repeat protein. Little is understood about its cellular role, however, a study using an immortalised mouse myoblast cell line found that the encoded protein has a role in myogenic differentiation[185]. The variant is of interest due to its high pathogenicity score and the gene's potential role in myocyte maturation.

4.6 Family D

A list of variants of interest can be found in Table 4.5.

Chr	Gene	Protein	Role	Variant	AA change	dbSNP(147)	gnomAD AF	CADD
2	*NEB*	nebulin	assembly and length maintenance of muscle thin filaments[190]	NM_004543:exon136:c.18509C>T	S6170L	rs202191938	0.0002	25.4
11	*UCP3*	UCP3	mitochondrial oxidative phosphorylation uncoupling[191]	NM_003356:exon7:c.881A>G	Y294C	.	.	25.4
15	*KBTBD13*	kelch repeat and BTB domain-containing protein 13	adaptor of Cullin E3 ubquitin ligase complex[192, 167]	NM_001101362:exon1:c.349G>T	G117C	.	.	24.6

Table 4.5: Family D final variants of interest
'AA' is amino acid. 'AF' is allele frequency. 'CADD' is the reported CADD score for the variant. Variant descriptions and amino acid change are the ones first reported by ANNOVAR and are based on the Refseq gene accession given, which is generally, but not necessarily, the canonical isoform.

4.6.1 *NEB* S6170L SNV

An *NEB* variant predicted to lead to a serine to leucine amino acid substitution at position 6170 (Refseq accession: NM_004543) was identified in family D. It is present at a low AF (gnomAD AF: 0.0002) and is predicted to be pathogenic (CADD score: 25.4).

NEB (OMIM accession: 161650) encodes nebulin, a very large protein exclusively expressed in skeletal muscle, responsible for correct assembly and length maintenance of thin filaments in striated muscle[190]. Variants in *NEB*, most often recessive, are associated with nemaline myopathy (NM), a heterogeneous genetic disease causing muscle weakness and hypotonia[193]. NM presents histologically as malformed Z-discs and aggregates of Z-disc proteins[193]. Although there is no evidence that NM patients are susceptible to MH[194], the variant is of interest due to its high predicted pathogenicity and because of the protein's importance in the correct functioning of muscle. The *NEB* S6170L variant is reported as being of unknown significance in relation to NM on ClinVar (ClinVar accession: VCV000095119.1).

4.6.2 *KBTBD13* G117C variant

A SNV within *KBTBD13*, encoding KBTBD13, predicted to cause a glycine to cysteine substitution at amino acid position 117 (Refseq accession: NM_001101362), was identified in individual BW1 of family D. Two out of four reads at the variant position give evidence of the variant's presence. There were no reads mapping to this variant position for the other individuals of family D and therefore their genotype for the variant is unknown.

Although there is low evidence for the existence of this variant, and no evidence of segregation due to lack of coverage, this variant is reported due to KBTBD13's association with muscle disease: variants in *KBTBD13* have been associated with nemaline myopathy[195, 196]. KBTBD13 is an adaptor of the Cullin E3 ubquitin ligase complex and a suggested pathogenesis of *KBTBD13*-mediated nemaline myopathy is that pathogenic variants of *KBTBD13* prevent proper turnover of important muscle-specific proteins[192, 167]. Additionally, the suspected variant in family D is predicted to be pathogenic (CADD score: 24.6). This variant is of interest due to the gene's association with muscle disease.

4.6.3 *UCP3* Y294C SNV

A *UCP3* SNV, predicted to lead to a tyrosine to cysteine substitution at amino acid position 294 (Refseq accession: NM_003356), was identified in family D. The variant appears to be novel as it was not found on any database. It is predicted to be pathogenic (CADD score: 24.6).

UCP3 encodes mitochondrial uncoupling protein 3 (UCP3), belonging to a family of transmembrane channel proteins that localise to the inner mitochondrial membrane[191]. *UCP3* is highly and specifically expressed in skeletal muscle[197]. As the name suggests, uncoupling proteins are understood to uncouple the electron transport chain and ADP phosphorylation by allowing passive proton leak down the proton gradient[191]. The phosphorylation state (specifically phosphorylation of tyrosine and serine) of UCP3 may be a mediator of its propensity for proton passage[76]. UCP3 is thought to play a role in MDMA-induced hyperthermia[74], which has symptoms similar to that of an MH episode (see Chapter 1.2.4).

Segregation analysis was carried out on this *UCP3* variant due to the gene's role in skeletal muscle energy metabolism and its association with a disorder of similar aetiology to MH.

4.6.3.1 Segregation analysis

Sanger sequencing was carried out on individuals 2040 (MHS, heterozygous for *CACNA1S* T1009K), 2041 (MHN) and 2013 (MHS, not a carrier of *CACNA1S* T1009K). The variant was not carried by any of these individuals (Figure 4.9), and the variant was therefore excluded from further analyses.

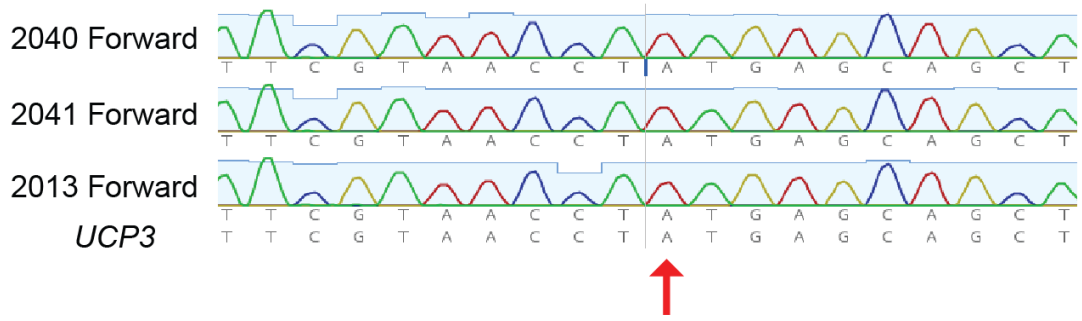

Figure 4.9: Alignment of Sanger sequencing results for individuals 2040, 2041 and 2013 against the *UCP3* nucleotide sequence
The *UCP3* Y294C SNV position is indicated by the red arrow. The results indicate that these individuals do not carry the variant.

4.7 Family E

A list of variants of interest can be found in Table 4.6.

Chr	Gene	Protein	Role	Variant	AA change	dbSNP(147)	gnomAD AF	CADD
1	*TMOD4*	tropomodulin 4	control of actin filament length[198]	NM_013353:exon9:c.994A>G	M332V	rs143662204	0.0045	13.84
6	*SLC16A10*	monocarboxylate transporter 10	skeletal muscle plasma membrane aromatic amino acid transporter[199]	NM_018593:exon3:c.859A>G	I287V	rs116322832	0.0015	1.32
6	*SYNE1*	nesprin-1	linking of nucleoskeleton and cytoskeleton[200]	NM_033071:exon116:c.21164A>G	K7055R	rs145882956	0.0002	0.468
8	*DUSP4*	dual specificity protein phosphatase 4	possible role in muscle cell differentiation[122]	NM_001394:exon1:c.82G>A	A28T	rs113228022	0.001	24.6

Table 4.6: Family E final variants of interest
'AA' is amino acid. 'AF' is allele frequency. 'CADD' is the reported CADD score for the variant. Variant descriptions and amino acid change are the ones first reported by ANNOVAR and are based on the Refseq gene accession given, which is generally, but not necessarily, the canonical isoform.

4.7.1 *TMOD4* M332V SNV

A *TMOD4* SNV predicted to lead to a methionine to valine substitution at amino acid position 332 (Refseq accession: NM_013353) was identified in family E. The variant has a gnomAD AF of 0.0045, which may be considered too high to be pathogenic for MH when considering its incidence rate. The variant has mixed pathogenicity predictions (CADD score: 13.8).

TMOD4 encodes tropomodulin 4, an actin regulatory protein expressed highly and specifically in skeletal muscle which controls actin filament length by capping the pointed end of an actin filament to prevent further polymerisation and depolymerisation[198]. The variant is of interest due to the genes role in skeletal muscle contraction.

4.7.2 *SLC16A10* I287V SNV

A *SLC16A10* SNV predicted to lead to an isoleucine to valine substitution at amino acid position 287 (Refseq accession: NM_018593) was identified in family E. The variant identified is not predicted to be pathogenic (CADD score: 1.3).

SLC16A10 encodes monocarboxylate transporter 10, a plasma membrane Na^+-dependent transporter which controls extracellular aromatic amino acid homeostasis by the cellular efflux of amino acids[199]. The gene is highly expressed in skeletal muscle[199]. It has not been previously linked to disease. The variant is reported here due to the gene's high skeletal muscle expression, however it is unclear how the gene's role could have a relationship to the pathogenesis of MH.

4.7.3 *SYNE1* K7055R SNV

A SNV within *SYNE1*, predicted to lead to a lysine to arginine substitution at amino acid position 7055 (Refseq accession: NM_033071), was identified in family E. The variant is not predicted to be pathogenic, with a CADD score of 0.5.

SYNE1 encodes nesprin-1, a large nuclear membrane-associated protein which makes up part of a protein complex linking the nucleoskeleton and the cytoskeleton[200]. Variants in *SYNE1* have been associated with Emery-Dreifuss muscular dystrophy (EDMD), a disease most commonly characterised by muscle contractures, elevated CK, myotonia and limb-girdle muscular dystrophy[200]. Noteably,

there is significant overlap between symptoms of EDMD and myotonic dystrophy, the diagnosis present within family E. Additionally, variants in *SYNE1* have been linked to arthrogryposis multiplex congenita, a non-progressive muscle disease caused by abnormal muscle tissue fibrosis, resulting in muscle shortening, which leads to joint contracture from birth[201]. The variant is of interest due to the gene's association with a muscle disease with symptoms closely related to the the muscle disorder carried by family E.

4.7.4 *DUSP4* A28T SNV

A *DUSP4* SNV, predicted to lead to an alanine to threonine substitution at amino acid position 28 (Refseq accession: NM_001394), was identified in family E. The variant is predicted to be pathogenic (CADD score: 24.6).

DUSP4 encodes dual-specificity phosphatase 4 (Dusp4). Recently, Dusp4 has been implicated in skeletal myocyte differentiation - the first study which has found a muscle cell-specific role for the protein[122]. The variant is of interest because it is plausible that a protein involved in myocyte differentiation could lead to incorrect functional expression of proteins involved in calcium homeostasis or ECC.

4.7.5 Other variants of interest

Two other variants of note identified in family E are discussed below.

4.7.5.1 Previously identified *RYR1* R1679H variant

This family has a previously identified RyR1 variant, R1679H. Although the carrier status of this variant for family E was known, its presence provided an opportunity to test whether the established bioinformatic pipeline was capable of identifying the variant, and retaining it through the filtering steps.

After processing of the WES data of family E through the bioinformatic pipeline, the variant was present in the final variants of interest list (not included in Table 4.6), confirming the capability of the pipeline to identify rare, predicted protein-changing variants which were segregating with MH within the family, despite some degree of discordance.

4.7.5.2 *RYR1* E2820Q variant

The WES data indicated that individual 1247 was a carrier of an *RYR1* SNV, NM_000540.2:c.8452G>T, predicted to lead to a glutamic acid to glutamine substitution at amino acid position 2820 (Refseq accession: NM_000540). The variant was predicted to be pathogenic and was not present on any database suggesting it to be novel.

Two out of eight reads suggested the presence of the variant, however, lack of coverage for the three other WES samples meant that their genotype could not be inferred (the variant was identified prior to the new batch of WES being carried out). Sanger sequencing was carried out, revealing that individual 1247 did not carry this variant (Appendix A.10). This suggested a WES error, although alternatively allele dropout may have occurred during both WES and Sanger sequencing. However, the variant was not pursued further.

4.8 Family F

A list of variants of interest can be found in Table 4.7.

Chr	Gene	Protein	Role	Variant	AA change	dbSNP(147)	gnomAD AF	CADD
11	*ANO1*	anoctamin 1	membrane depolarisation in smooth muscle by chloride transport [202]	NM_018043:exon15:c.1496T>G	V499G	rs775543205	3.97E-05	24.2
19	*SLC25A42*	mitochondrial coenzyme A transporter SLC25A42	mitochondrial transport of coenzyme A [203]	NM_001321544:exon2:c.26C>T	P9L	rs117940121	0.0009	12.15

Table 4.7: Family F final variants of interest
'AA' is amino acid. 'AF' is allele frequency. 'CADD' is the reported CADD score for the variant. Variant descriptions and amino acid change are the ones first reported by ANNOVAR and are based on the Refseq gene accession given, which is generally, but not necessarily, the canonical isoform.

4.8.1 *ANO1* V499G SNV

An *ANO1* SNV, predicted to lead to a valine to glycine substitution at amino acid position 499 (Refseq accession: NM_018043), was identified in family F. The variant is predicted to be pathogenic (CADD score: 24.2).

ANO1 encodes anoctamin 1 (also known as TMEM16A), a calcium-activated chloride channel (CaCC). Anoctamin 1 is important for contraction of smooth-muscle: channel opening leads to an efflux of Cl^-, causing membrane depolarisation and subsequently muscle contraction[202]. Another CaCC, anoctamin 5 (encoded by *ANO5*), is a homologue of anoctamin 1 which is primarily expressed in skeletal muscle[204]. *ANO5* variants are associated with Miyoshi muscular dystrophy type 3 and limb girdle muscular dystrophy type 2L[205]. Recent evidence by knock-down of anoctamin 5 suggests that it is responsible for the correct functional expression of RYR1, DHPR and SERCA proteins[206]. Although anoctamin 1 is not known to play a role in skeletal muscle in humans, it was recently reported that ANO1 contributes to skeletal muscle contraction intensity by action potential modulation in zebrafish[207]. The variant is of interest due to the gene's potential role in action potential modulation: an MH episode involves persistent activation of the RyR1 channel which could be caused by dysfunction of molecular machinery involved in action potential generation.

4.8.2 *SLC25A42* P9L SNV

A SNV within *SLC25A42* identified in family F is predicted to lead to a proline to leucine substitution at amino acid position 9 (Refseq accession: NM_001321544). The variant has mixed pathogenicity predictions (CADD score: 12.15).

SLC25A42 (OMIM accession: 610823) encodes a mitochondrial coenzyme A (CoA) transporter protein. CoA is a substrate in a number of crucial metabolic pathways including the citric acid cycle and fatty acid oxidation. The gene is highly expressed in skeletal muscle[121]. Variants in the gene are causative of mitochondrial myopathies[203]. The variant is of interest due to the gene's role in energy metabolism.

4.9 Variants of interest shared between families

4.9.1 *WASHC2C* variants

Rare, predicted pathogenic variants within the gene *WASHC2C* were identified in family A, B and F (Table 4.8). Two of these variants are presumed novel as they could not be found in any database. The *WASHC2C* variant harboured by family B has been previously reported and has a gnomAD AF of 0.00621. For family A, the WES data suggests that the variant completely segregates with MH susceptibility amongst the five individuals for which WES were obtained. For family B, the WES data suggests that the variant is heterozygous in MHS individual 1772 (MHS) (44 reads out of 99 with variant sequence), but not present in MHS individual 919 (MHS) (0 out of 35 reads with variant sequence). There is no coverage at the variant position for all other WES samples from family B. For family F, the WES data suggests that all three individuals (all MHS) harbour the *WASHC2C* variant.

WASHC2C encodes WASH complex subunit 2C, also known as FAM21C, FAM21A and VPEF. The protein is a component of the WASH complex, involved in endosomal sorting by regulation of the Arp2/3 complex, which is itself involved in the induction of actin polymerisation[208]. Another component of the WASH complex, strumpellin, encoded by *WASHC5* (OMIM: 610657), has been implicated in numerous human disorders. Variants in *WASHC5* have been associated with spastic paraplegia[209] (a disease of motor neurons), Ritscher-Schinzel syndrome[210] (a developmental disorder affecting function of a number of organs), and hyperkalemic periodic paralysis[211] (a muscle disorder). Additionally, variants in *WASHC4*, which encodes another member of the WASH complex, have been implicated in a disorder characterised by mental retardation[212]. The pathophysiology of these variants in relation to these disorders is not currently understood.

Although there is no clear link between *WASHC2C* and a muscle phenotype, it is interesting that variants affecting the WASH protein complex are implicated in a range of disease phenotypes. The fact that rare *WASHC2C* variants were identified in three of the six families under study is of interest. Further study, beginning with segregation analysis within each of the three families for the respective variants, could be undertaken to understand whether these variants may be implicated in MH-susceptibility.

Family	Variant	AA change	dbSNP(147)	gnomAD AF	CADD
A	NM_001169107:exon11:c.972C>A	D324E	.	.	27
B	NM_001367404:exon8:c.724G>A	A242T	rs201469599	0.00621	26.4
F	NM_001367404:exon19:c.1933G>C	G645R	.	.	23.6

Table 4.8: *WASHC2C* variants identified in families A, B and F
'AA' is amino acid. 'AF' is allele frequency. 'CADD' is the reported CADD score for the variant. Variant descriptions and amino acid change are the ones first reported by ANNOVAR and are based on the Refseq gene accession given, which is generally, but not necessarily, the canonical isoform.

4.9.2 *GCGR* R413W SNV

A rare, predicted pathogenic variant within *GCGR*, NM_000160:exon14: c.1237C>T (dbSNP: rs757894502), was identified in families A and B. The variant is predicted to lead to an arginine to tryptophan substitution at amino acid position 413 (Refseq accession: NM_000160). The variant has a gnomAD AF of 0.00005. For family A, the WES data suggests that the variant completely segregates with MH susceptibility amongst the five individuals for which WES were obtained. For family B, the WES data suggests that the variant is homozygous in individuals 82 (MHS) and 456 (MHS), heterozygous in individuals 1772 (MHS), 1367 (MHS), 1412 (MHS), 26 (unknown) and 919 (MHS), and not present within individuals 27 (MHS_h), 251 (MHN), and 1731 (MHS). There is no coverage at the variant position for individual 533 (MHS).

GCGR (OMIM accession: 138033) encodes the glucagon receptor, a G-protein coupled receptor which binds glucagon and is responsible for the regulation of blood glucose levels and glucose homeostasis. The amino acid position of the variant (413) resides within the cytoplasmic domain of the protein[213]. The gene is expressed highly in hepatic tissue, but not in skeletal muscle tissue. Although glucagon signalling is not known to play a part in MH, it remains possible that dis-regulation of glucose levels may have an effect on skeletal muscle energy homeostasis. Alternatively, the fact that both family A and B carry this rare variant could be explained by an unknown familial link.

4.10 Copy-number variation detection

4.10.1 Detection of known polymorphism CNVs leading to method validation

A common polymorphism copy-number variant (CNV) that has been previously identified is reported in this section. This CNV is not thought to be of relevance to MH or muscle disease, but is reported to validate the viability of CNV detection using CNVkit.

4.10.1.1 Chromosome 11 duplication

A chromosome 11 duplication, approximately 30 kb in length, was identified in WES sample 2253 from family E. The duplication has been reported previously (dbVar accession: esv2759831) and has an estimated AF of 0.67 (out of 270 samples). Coverage of WES samples 2253 and 2259 over the chromosomal region containing the suspect CNV is plotted in Figure 4.10.

4.10.2 Family A CNVs

No CNVs of interest were identified in family A.

4.10.3 Family B CNVs

Two CNVs of interest were identified in family B.

4.10.3.1 Chromosome 16 duplication

Individual 1731 appeared to have a duplication (4-fold normal copy number) spanning approximately 106 kb, from base 74306327 to 74413250, at chromosomal location 16q23.1. CNVkit reports the copy number of the region to be 4, which which is likely from inspection of the coverage plot (Figure 4.11). This region harbours the gene *LOC283922* whose role has not previously been reported, however it is a pseudogene of *PDPR* (phosphate dehydrogenase pyruvate regulatory subunit), a key regulatory subunit of the pyruvate dehydrogenase complex. It is unclear whether or not the pseudogene is expressed.

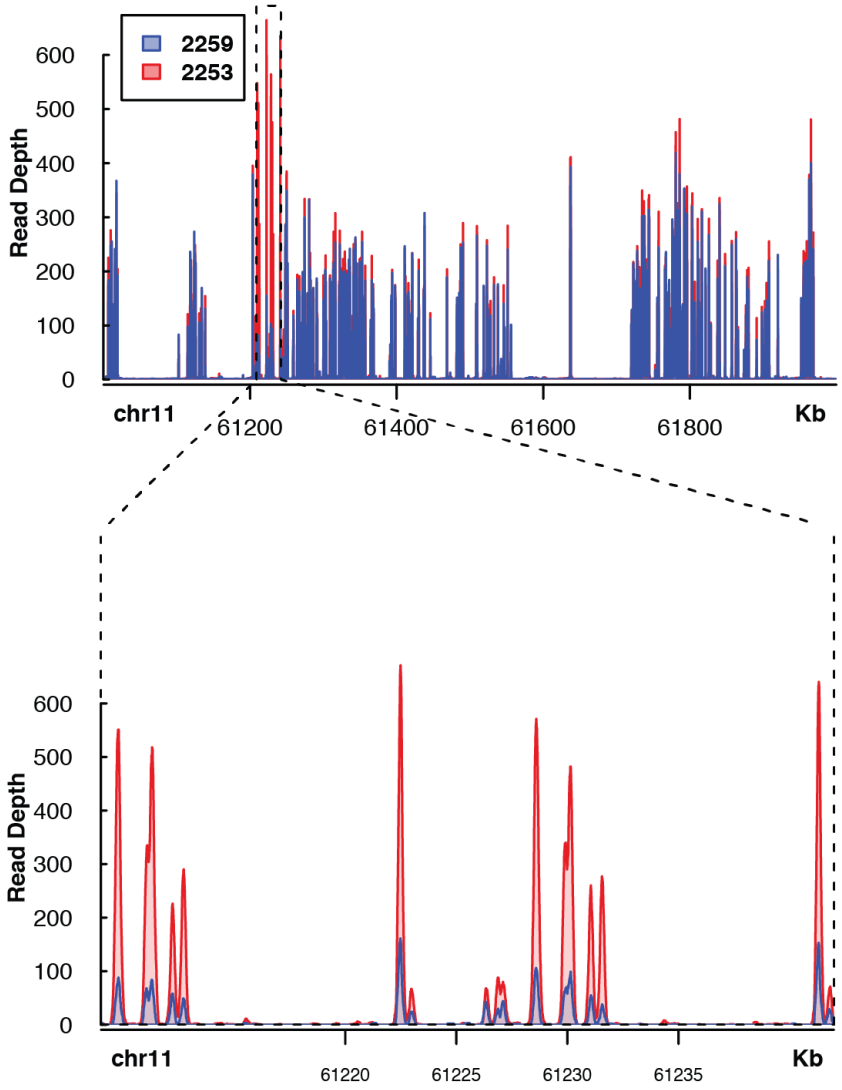

Figure 4.10: Evidence of a chromosomal duplication of chromosome 11 in WES sample 2253 from family E

The x-axis gives the chromosomal location, the y-axis the read depth.

4.10.3.2 Chromosome 17 deletion

An approximately 34 kb homozygous deletion within chromosome 17 (bases 64862900 - 64897300) was predicted by CNVkit for the WES sample from individual 1412. However, as reads were present upon inspection of this region, it is likely that there is at least one copy of the region present. This region harbours the gene *LRRC37A3*, whose role is not currently understood. Expression of the gene is reported as low on the Human Protein Atlas. Individual 1731 (MHS) did

not appear to harbour the deletion upon inspection of WES alignment data.

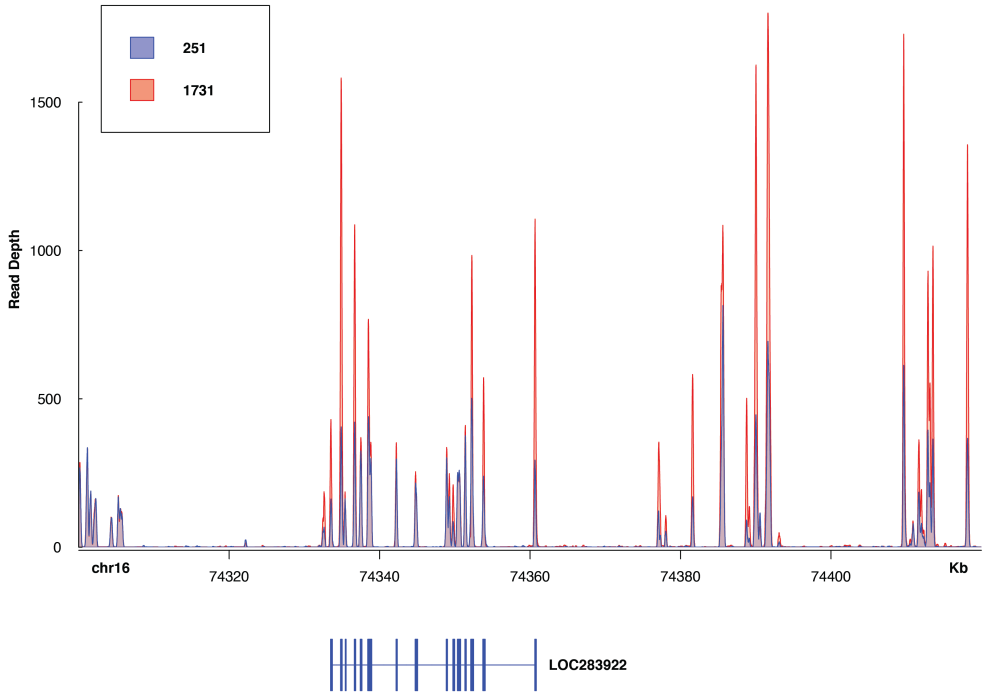

Figure 4.11: Evidence of a chromosomal duplication of chromosome 16 in WES sample 1731 from family B

The x-axis gives the chromosomal location, the y-axis the read depth. The position of the *LOC283922* gene is shown below the x-axis.

4.10.4 Family C CNVs

No CNVs of interest were identified within family C.

4.10.5 Family D CNVs

No CNVs of interest were identified within family D.

4.10.6 Family E CNVs

A chromosome 15 duplication of approximately 11 kb, from bases 101752800 to 101764500, was detected in the WES of individual 2253 (Figure 4.12). CNVkit reports the copy number of the region to be 6. The region is not known to harbour any genes.

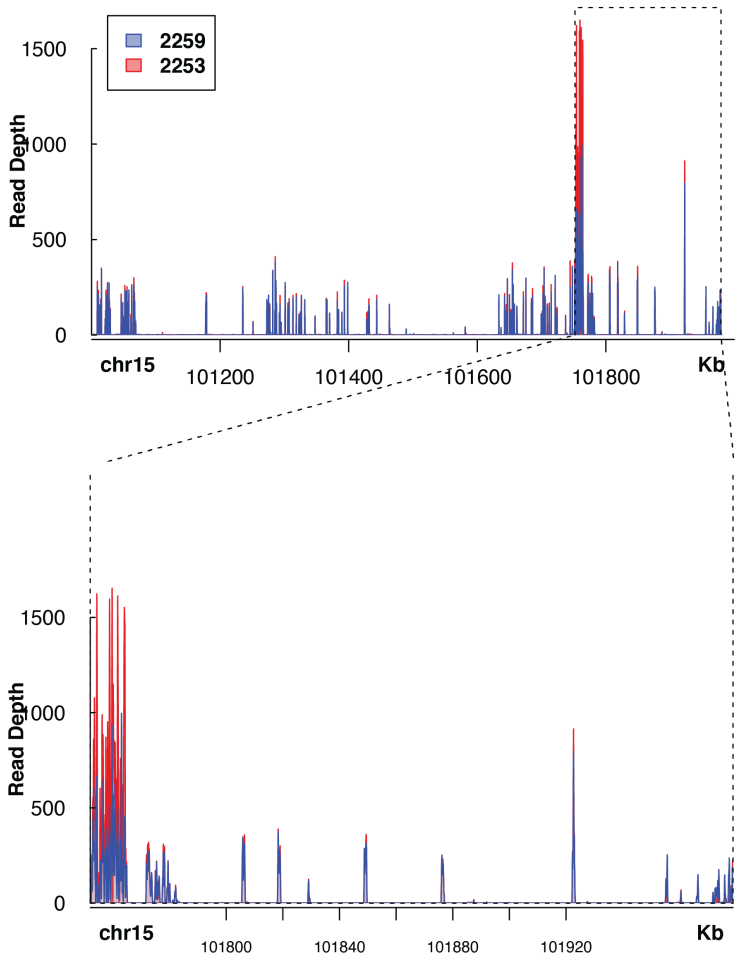

Figure 4.12: Evidence of a chromosomal duplication of chromosome 15 in WES sample 2253 from family E

The x-axis gives the chromosomal location, the y-axis the read depth.

4.10.7 Family F CNVs

No CNVs were detected within family C.

4.10.8 Shared CNVs between families

4.10.8.1 Chromosome 22 duplication in family B and E

A duplication of chromosome 22 was detected in the MHS individuals under CNV investigation in family B and E (Figure 4.13 & 4.14). The duplication spans approximately 60 kb, from base 42500000 to 42560000, at chromosomal location 22q13.2. The suspected duplicated region contains the genes *SERHL*

and *RRP7A*.

This CNV has been reported in various publications[214,215,216]. Artuso et al.[214] reports the duplication size as 49 kb between bases 42511781-42561110. Shu et al.[216] reports the size as 42 kb between bases 42508600-42550874.

Shu et al.[216] reported an AF for a copy number of 3 and 4 to be 0.20 and 0.02, respectively, amongst 3913 individuals tested.

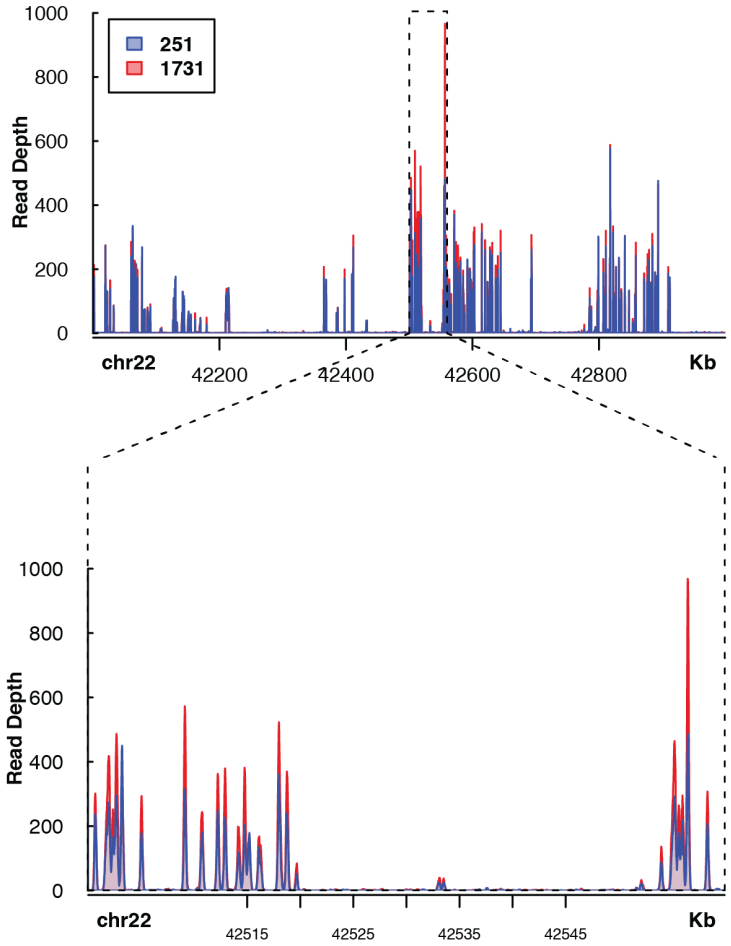

Figure 4.13: Evidence of a chromosomal duplication of chromosome 22 in WES sample 1731 from family B

The x-axis gives the chromosomal location, the y-axis the read depth.

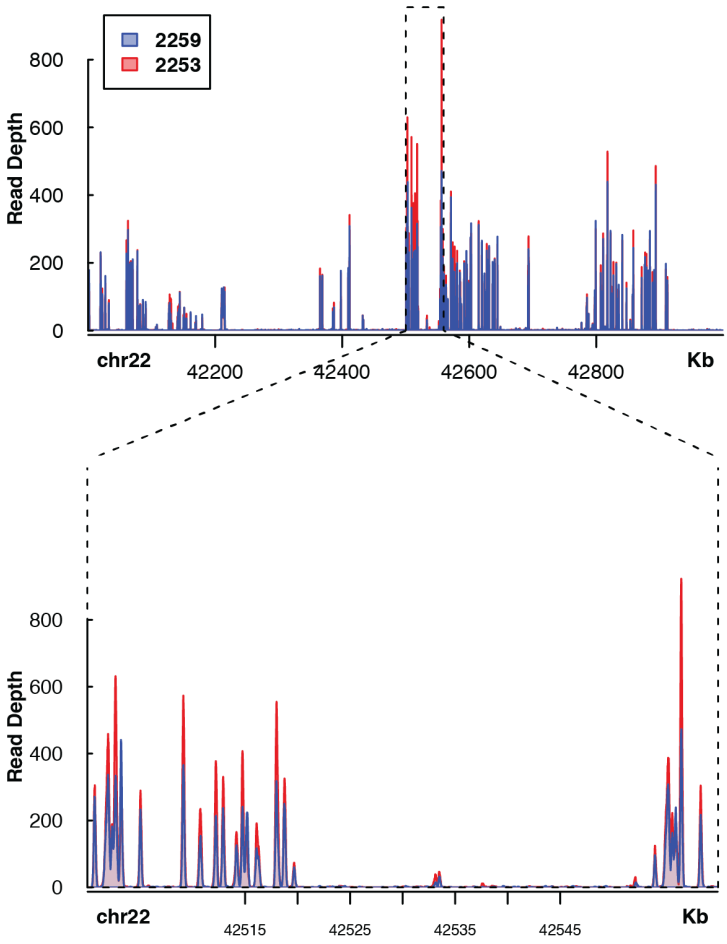

Figure 4.14: Evidence of a chromosomal duplication of chromosome 22 in WES sample 2253 from family E

The x-axis gives the chromosomal location, the y-axis the read depth.

4.10.9 *KCNJ* gene family

The *KCNJ* genes encode a family of inward rectifier potassium channel (Kir) proteins which function to control resting membrane potential within excitable cells and repolarisation of action potential by the conductance of K^+ into the cell[217]. Kir2 isoforms, including Kir2.2 and Kir2.1 (encoded by *KCNJ12* and *KCNJ2*), are the main Kir subfamily expressed in mammalian skeletal muscle[121]. The channels form both homo- and hetero- tetrameric assemblies[217]. Identifying variation in these proteins may be of interest to MH research as they play an

important role in excitability of myocytes and thus it is plausible that their dysfunction could lead to abberant action potential generation and ECC. Additionally, *KCNJ2* falls within the MHS2 locus, with *KCNJ12* falling just outside of this locus.

The genes *KCNJ12* and *KCNJ18* are paralogous[218]. Due to the high sequence similarity of these paralogs, it is difficult to identify variants in these genes due to misalignment of sequence. This problem was originally identified because rare, predicted pathogenic variants which appeared in variants of interest lists were being detected within *KCNJ12* in all families under study. Inspection of the sequence alignments of the *KCNJ12* gene indicated that these were not true variants as there was significant variation between the read sequences and the genomic sequence to which they were aligning (Figure 4.15). Additionally, reads were not mapping to *KCNJ18*.

A group of researchers trying to identify variants within *KCNJ18* experienced a similar issue[218]. The researchers were trying to identify variants within *KCNJ18* as this gene is a candidate for the pathogenesis of thyrotoxic period paralysis: the gene falls within the disease locus and its molecular function corresponds to the condition's manifestation which is muscle weakness and transient hypokalemia (low blood potassium)[218]. The researchers found misalignment of these genes prevented identification of potential variants of interest and so developed a method to correctly map the genes[218]. Unfortunately, the paper's methodology was not explained sufficiently to allow it to be reproduced and the researchers could not be contacted.

It could be of interest to pursue sequencing of these genes not only due to their involvement in myocyte electrical excitability but also because it is likely that mapping issues have previously prevented the identification of pathogenic variants. Due to these mapping issues, it would be necessary to carry out Sanger sequencing on the genes using well-designed primers which anneal to regions that lack sequence similarity between the paralogs to identify variants.

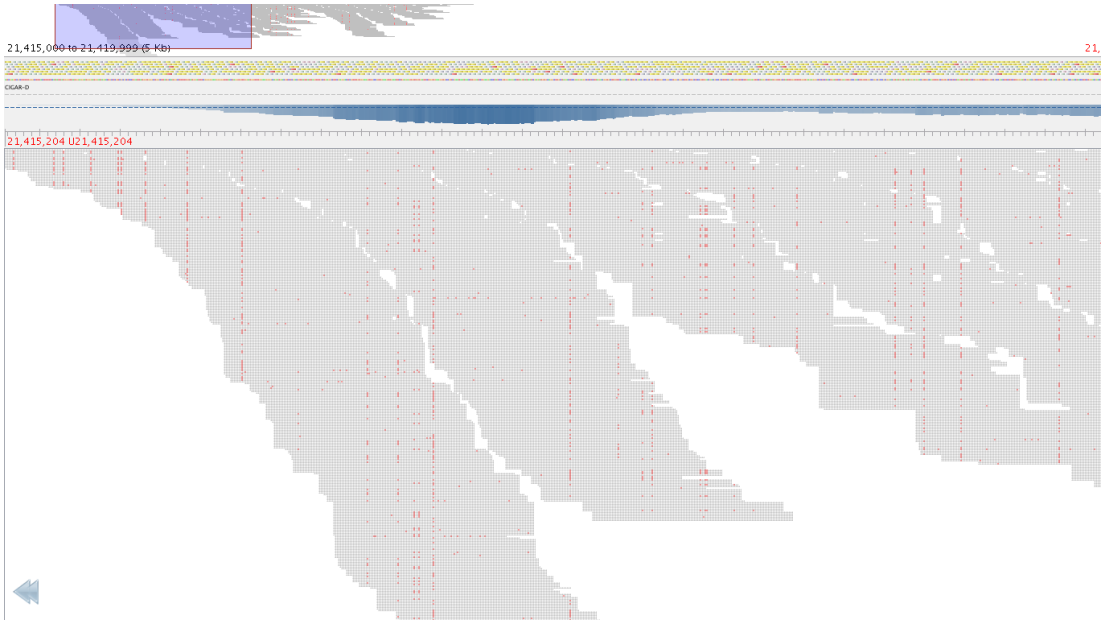

Figure 4.15: Visualisation of alignment of reads to the *KCNJ12* gene representative of all WES samples

WES sample 1652 is shown here at chromosomal position 21,415,000 - 21,420,000 of chromosome 17, which is the position of exon 3 of the *KCNJ12* gene (RefSeq accession: NM_021012.5). The red squares represent variation in the reads in comparison to the hg38 sequence which the read is aligned to. The alignment data suggests that the reads are mismapping to this region. This alignment represents alignment of all other WES samples analysed.

4.11 Summary

To summarise, variants of interest were identified in all six families under study. A number of variants were selected for segregation analysis. This ruled out all but one variant as segregating with MH-susceptibility. In family A, a SNV within the gene *TLR3* (predicted to lead to an G320R amino acid substitution) was revealed to be segregating with MH-susceptibility for all seven family members for which DNA was available. Due to the small sample size, further evidence should be gathered, such as a population MHN screen, before the variant is further pursued.

A number of rare, predicted pathogenic variants within the gene *WASHC2C* were identified. Although it is unknown how the function of the *WASH2C2*'s protein product could be related to the pathophysiology of MH, the variants were reported due to the fact that they were identified in three different families under study. It is unclear whether the *WASHC2C* variant identified in family B is segregating with MH-susceptibility as there is no coverage at the variant position for many of the WES samples. This is a good example of the power of the established bioinformatic pipeline to identify variants that may be of interest to MH, as this variant within family B would have been missed in other variant of interest identification pipelines. Additionally, a rare, predicted pathogenic variant was identified in the gene *GCGR*. It is unlikely that this variant is of interest to MH as the gene is not known to be expressed in skeletal muscle, however it was reported as the variant was identified in both family A and family B, who are not thought to be related.

CNV analysis identified predicted CNVs in the families. It is unknown whether or not these CNVs segregate with MH-susceptibility, as CNV analysis was only undertaken in one MHN and MHS individual within each family. The next step would be to analyse the other available WES for CNVs to ascertain whether there is segregation with the disorder for any of the identified CNVs. A previously reported chromosome 22 duplication appears to be present in family B and family E. Whether or not this is of interest is dependent on the exact copy number in each family as a single duplication of the region seems to be a common polymorphism, whereas a double duplication is significantly less common.

Mappability issues for two *KCNJ* genes were identified which are likely to inhibit identification of variation in these genes. It could be of interest to sequence these genes to identify potential variation.

Chapter 5

JSRP1 polymorphism association study

5.1 Introduction

The *JSRP1* gene encodes JP-45, a small 331 amino acid transmembrane protein, enriched in skeletal muscle tissue, which resides in the sarcoplasmic reticulum membrane[6]. It is thought to be a direct modulator of $Ca_v1.1$ activity[6]. Its N-terminal tail interacts with calsequestrin in the sarcoplasmic reticulum lumen and its C-terminal end interacts with $Ca_v1.1$ in the myoplasm[219]. JP-45 is required for the functional expression of $Ca_v1.1$, making it a key protein in the development and maintenance of healthy adult skeletal muscle[220].

A *JSRP1* NC_000019.10:g.2252991C>G variant (dbSNP: rs80043033) leading to a JP-45 glycine to alanine substitution at amino acid position 150 (G150A) was identified in family B. This glycine resides in the sarcoplasmic reticulum lumenal domain of JP-45[6]. The gnomAD AF of the variant is 0.2, classing it has a common polymorphism. Although both the *JSRP1* and *RYR1* genes reside on chromosome 19, *JSRP1* G150A and *RYR1* T4826I (also harboured by family B) variants are not genetically linked due to the distance between the genes (cytogenetic position 19p13.3 for *JSRP1* and 19q13.2 for *RYR1*). *In vitro* human cell line calcium release experiments have previously established that *JSRP1* G150A may alter the strength of excitation-contraction coupling in human skeletal muscle by leading to a decrease in DHPR activation sensitivity[221]. Yasuda et. al (2013)[221], with a sample size of 140 individuals, did not find a correlation

between susceptibility to malignant hyperthermia (MH) and *JSRP1* G150A variant status. Therefore the variant is not believed to directly cause - nor provide a protective role - from MH susceptibility. However, its ability to alter the extent to which the muscle cell reacts to an excitation event may mean that it can alter the expressivity of an MH-related variant, potentially classing it as a disease-modifying variant. This may explain some extent of the variability seen between individuals in the IVCT within family B, including both those with and without the *RYR1* T4826I variant. The discovery of this variant in family B provided an opportunity to carry out a robust genotype-phenotype association study due to the substantial number of DNA samples and IVCT results available.

The hypothesis for this association study is that individuals carrying *JSRP1* G150A will have reduced contracture values compared to individuals who do not carry the variant. Additionally, reduction in contracture values will be more pronounced for those homozygous for the variant.

5.2 Methods and method development

Extensive method development was required to determine a suitable assay for large-scale genotyping of *JSRP1* G150A. The following sections describe three types of assays that were attempted: high resolution melting (HRM) analysis, hybridisation probe analysis, and restriction length polymorphism (RFLP) genotyping. HRM and hybridisation probe assays were not adopted for genotyping due to a number of problems encountered. An RFLP assay was ultimately chosen for genotyping.

General methods used in this study can be found in Chapter 2. Primer and probe sequences, and primer sets, can be found in Appendix A.1, Appendix A.2, and Appendix A.3.

5.2.1 Confirming genotypes for use as controls

Sanger sequencing was carried out on family B gDNA samples 533, 919 and 1772 which confirmed the genotypes suspected through WES analysis (Appendix A.11): 533 did not carry the variant (wild-type), 919 was homozygous for the variant, and 1772 was heterozygous for the variant. These three samples were used as controls in method development and final genotyping.

5.2.2 High resolution melting assay

HRM assays are often the most time-effective and cost-appropriate method for large-scale genotyping, and so this method was initially chosen for genotyping of the *JSRP1* G150A variant. Using HRM primer set 2, an initial HRM optimisation assay using control samples 533 (wild-type) and 1772 (heterozygous) was able to successfully distinguish the two genotypes (Appendix A.12). However, when this assay was repeated with other samples, including homozygous 919, there was difficulty in differentiating wild-type from homozygous samples. Although an expected shift in the homozygous single melting peak with regards to the wild-type single melting peak was observed, the shift was not pronounced enough to easily distinguish between wild-type and homozygous genotypes (Appendix A.13). This issue was further compounded by the fact that many gDNA samples to be genotyped were older (some gDNA was isolated over 20 years ago) and so were likely to be degraded to some extent. Degraded gDNA can lead to melting curve shifts and can cause secondary peak formation even when no SNP is present within an amplicon, meaning that unless a significant heteroduplex peak is expected in the assay, some samples may be erroneously genotyped as heterozygous.

This assay was repeated with replacement of 0.5 μL H_2O with 0.5 μL 5x FastStart^{TM} Taq DNA Polymerase GC-RICH solution (Roche), but the addition of this additive did not seem to improve the assay (Appendix A.14).

5.2.2.1 HRM primer redesign

HRM primers were redesigned to produce an amplicon shorter than that produced by HRM primer set 2 (amplicon length of 106 bp). Shorter amplicon sizes allow greater resolution of SNVs by HRM. With a combination of newly designed primers and primers from the original HRM primer sets, four new primer sets were tested by PCR for amplification of the target product. HRM primer set 4 and HRM primer set 6, designed to produce amplicons of length 49 and 44 bp, respectively, appeared to produce a product of the expected size over two annealing temperatures tested. HRM primer set 5 produced an amplicon of the expected size (88 bp). HRM primer set 3 did not produce a product. HRM primer sets 4 and 6 were chosen going forward as they produced the smallest amplicon sizes.

An initial HRM assay with HRM primer set 6 using three control genotypes produced the expected melting peaks, with a more pronounced secondary peak

present for the heterozygous control (Figure 5.1) when compared to assays with HRM primer set 2. However, there was no improvement in the degree of shift of the homozygous control single peak relative to the wild-type single peak. The assay was repeated with HRM primer set 4, which gave similar results.

A method for resolving homozygous genotypes from wild-type genotypes using this HRM assay was devised which involved mixing suspected homozygous genotype reactions with known wild-type genotype reactions (and vice-versa). It was theorised that mixing equal amounts of wild-type and homozygous products together should result in a melting curve which matches a heterozygous melting curve due to the formation of heteroduplexes. This theory was tested by mixing 5 μL of wild-type 533 HRM reaction mixture with 5 μL homozygous 919 HRM reaction mixture. A melting curve-only programme was carried out on this mixed reaction. A heterozygous double melting peak which matched that of a known heterozygous (529) melting peak was observed (Figure 5.2). This demonstrated that it was possible to use this approach to distinguish homozygous genotypes from wild-type genotypes using HRM.

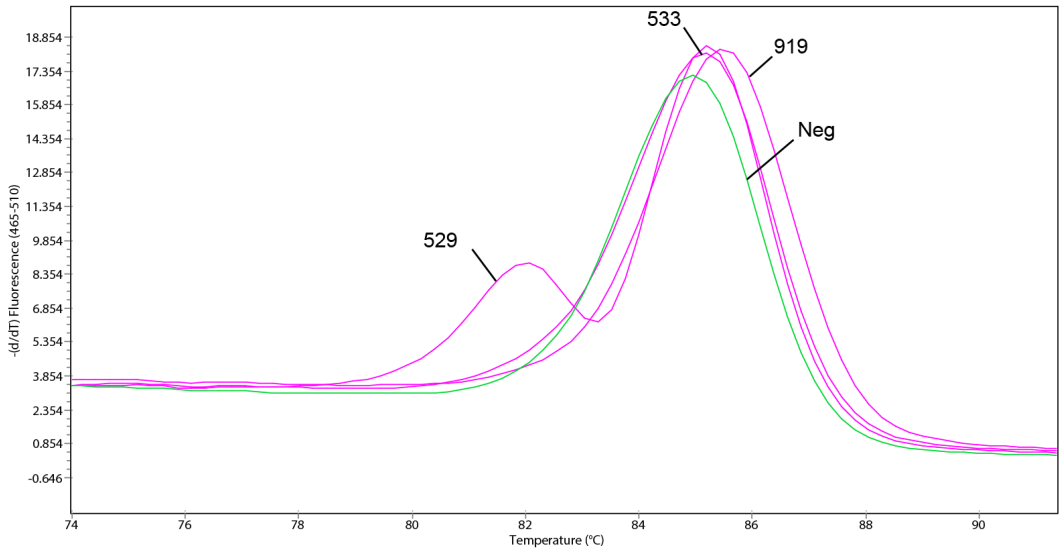

Figure 5.1: Representative *JSRP1* G150A genotype melting curve with HRM primer set 6

Heterozygous 529 shows a characteristic double peak. Wild-type 533 shows a characteristic single peak. Homozygous 919 shows a characteristic single peak shifted from the wild-type peak. Negative control amplification was observed, with a melting peak which matched that of the wild-type melting peak.

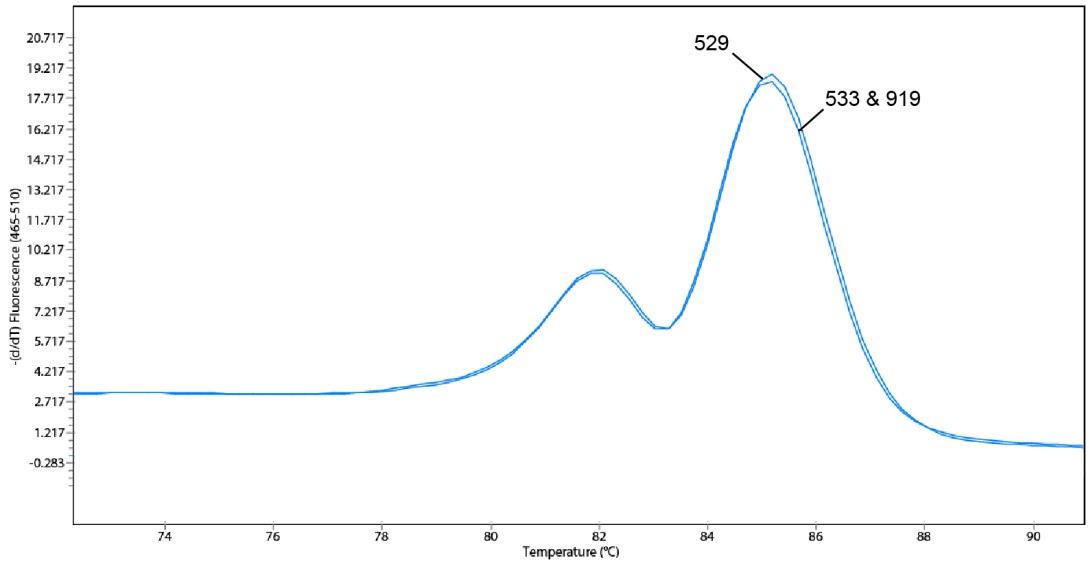

Figure 5.2: Initial mixed melt curve for *JSRP1* G150A homozygous genotype confirmation

Mixing of equal parts of the HRM reaction mixture for wild-type sample 533 and homozygous sample 919 produced a double peak closely resembling the heterozygous 529 melting curve.

5.2.2.2 Negative control amplification

Negative control amplification was observed in most assays carried out using primer set 4 and 6, and this negative control melting curve was similar to that of the wild-type melting curve (Figure 5.1). DNA contamination, or primer dimer formation, were two potential explanations for the negative control amplification. As amplification was still observed after changing stock solutions multiple times, including the LightCycler® 480 High Resolution Melting Master (Roche) and primer stocks, the amplification was likely caused by primer dimers. Testing this hypothesis would require the determination of the length of the negative control amplification product. A length that differed from the length of the target product (44 bp) would suggest primer dimer formation.

5.2.2.3 Native polyacrylamide gel electrophoresis for determination of negative control amplification size

Non-denaturing polyacrylamide gel electrophoresis was carried out to resolve possible differences in the size of the negative control amplification product and the positive control products. A 20% polyacrylamide gel was prepared containing the following: 6.25 mL 40% 29:1 acrylamide/bisacrylamide, 1.25 mL 5x TBE

buffer, 5 mL dH_2O, 125 μL 10% (w/v) APS (ammonium persulfate) and 12.5 μL TEMED (tetramethylethylenediamine). The electrophoresis was carried out at 120 V for 8 hours. One μL GeneRuler Ultra Low Range DNA Ladder (ThermoFisher, Cat. No. SM1213) was loaded alongside products for fragment size estimation.

Analysis of the product sizes indicated that the size of the positive control amplification products (44 bp) and negative control amplification product were the same (Figure 5.3), suggesting that presence of DNA contamination was the cause of negative control amplification.

Because of the continued issues with development of an HRM assay, an alternative assay for large-scale genotyping was attempted.

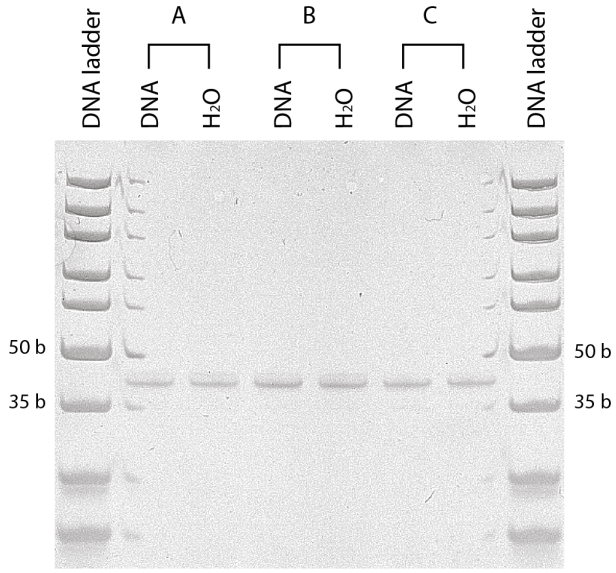

Figure 5.3: 20% native polyacrylamide gel for determination of negative control amplification product size for the *JSRP1* HRM assay

DNA ladder is the GeneRuler Ultra Low Range DNA Ladder (ThermoFisher). Reaction A contained 0.4 μL of 5 μM primers with no additives. Reaction B contained 0.6 μL of 5 μM primers with 0.5 μL 5x FastStart^{TM} Taq DNA Polymerase GC-RICH solution (Roche). Reaction C contained 0.4 μL of 5 μM primers and 0.5 μL GC-RICH solution. gDNA from individual 1731 was added to reactions labelled 'DNA'. dH_2O was used in place of DNA for reactions labelled 'H_2O'.

5.2.3 Hybridisation probe assay

5.2.3.1 Principle of the hybridisation probe assay

Hybridisation probe (HybProbe) analysis is a genotyping technique which involves the use of fluorescently-labelled nucleic acid hybridisation probes for allele detection and discrimination. It is similar to HRM in that discrimination of alleles is achieved by detection of changes in fluorescence - caused by DNA strand melting - over a temperature range (melting curve differences). A HybProbe assay requires typical forward and reverse primers for amplification of a short (∼150 bp) amplicon containing the allele of interest, as well as *anchor* and *sensor* probes which are designed to anneal adjacently to the same DNA strand within the amplicon region. A detailed explanation of the principle of the HybProbe assay is given in Figure 5.4. A benefit of doing a HybProbe assay is that, unlike HRM, off-target amplification should not alter the melting curve due to the requirement for sequence complementarity between the probes and the target amplification product. For the same reason, primer dimers are generally not an issue. Ultimately, a HybProbe assay tends to allow clear discrimination of wild-type, heterozygous and homozygous carriers for a given SNP.

In this study, a fluorescein-labelled sensor probe and a Cy5-labelled anchor probe were used. Fluorescein has an absorbance range of ∼430-520 nm and an emission range of ∼500-600 nm, whereas Cy5 (the acceptor) has an absorbance range of ∼530-675 nm and an emission range of ∼630-780 nm. This spectral overlap, demonstrated in Figure 5.5, allows FRET to occur when fluorescein is exposed to its excitation wavelength.

5.2.3.2 Probe and primer design

The LightCycler® Probe Design Software 2.0 (Roche) was used to design the HybProbe assay primers and probes. Primer and probe sequences can be found in Appendix A.1 and Appendix A.2, respectively. The sensor probe, 16 nucleotides in length, was designed, by default, to complement the sense strand of the variant sequence. The primers designed by the software (HybProbe primer set 1) were tested by PCR over a range of temperatures but were not used due to off-target amplification (Appendix A.15). The fact that these primers were short, both 18 nucleotides, likely contributed to these problems as shorter primer lengths

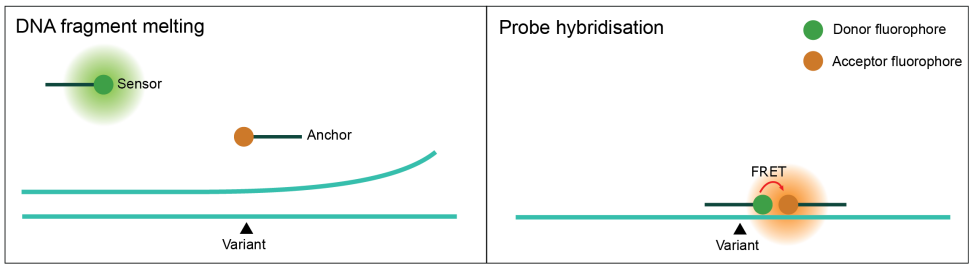

Figure 5.4: Principle of the HybProbe assay
The sensor probe, which has a fluorophore ligated to its 3' end, is designed to hybridise to the sequence containing the SNP, and is able to tolerate single nucleotide mismatches. The anchor probe is designed to hybridise to the 3' side of the sensor probe, on the same DNA strand, and has a different fluorophore ligated to its 5' end. Additionally, the anchor probe contains a phosphate group at its 3' end to prevent DNA elongation. One fluorophore is the *donor*, the other fluorophore the *acceptor*. Light of a specific wavelength excites the donor fluorophore, which will then transfer this energy to the acceptor fluorophore in a process known as Förster resonance energy transfer (FRET). FRET will occur only if the donor and acceptor fluorophores are within a range of a few bp, with FRET efficiency dropping considerably with distances of greater than ~10 nm[222]. Excitation of the acceptor fluorophore will then lead to detectable emission at a specific wavelength. Over a certain temperature range, the anchor and sensor probes will dissociate from the DNA strand, causing a decrease in FRET. The temperature at which the sensor probe dissociates is dependent on the sequence complementarity, allowing detection of a variant.

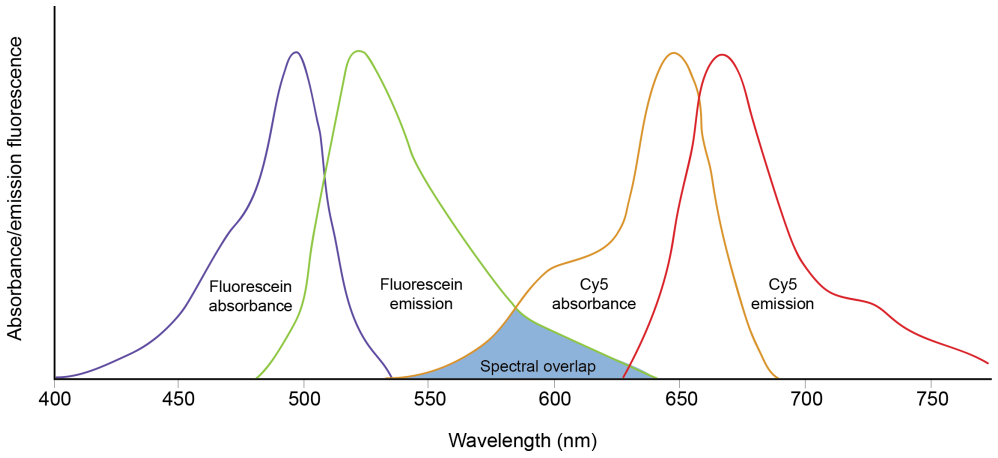

Figure 5.5: Absorption and emission spectra of fluorescein and Cy5 fluorophores
The spectral overlap represents the wavelengths at which absorption by Cy5 of emitted energy from fluorescein occurs. The y-axis represents relative fluorescence intensity (no units).

increase the possibility of off-target amplification. The primers were redesigned (HybProbe primer set 2) and were determined to exclusively amplify a product

of an expected size of 183 nucleotides by gel electrophoresis.

5.2.3.3 HybProbe assay method development

Initially, the assay was carried out using PerfeCTa® qPCR ToughMix® (QuantaBio) to the manufacturer's suggested protocol. Unless otherwise stated, each 10 μL reaction contained the following: 0.5 μL each of 5 μM forward and reverse primers, 100 nM each of sensor and anchor probes, 5 μL 2x PerfeCTa® qPCR ToughMix® (QuantaBio), 0.5 μL 50 ng μL^{-1} gDNA, and a variable volume of dH_2O to make up to 10 μL. LightCycler® machine conditions were varied as stated but were initially as follows: 95 ℃ for five minutes for initial denaturation; 40 cycles of 95 ℃ for 10 seconds, 60 ℃ for 20 seconds, and 72 ℃ for 10 seconds for amplification; 95 ℃ for one minute, 60 ℃ for three minutes, and a target of 90 ℃ for melting curve fluorescence acquisition; and lastly 45 ℃ for 30 seconds for cooling. A single fluorescence acquisition was taken upon reaching the 60 ℃ target during each amplification cycle. Five (continuous) fluorescence acquisitions per ℃ were taken as the temperature was increased at a ramp rate of 0.11 ℃ per second between 45 ℃ and 90 ℃ during the melt curve cycle.

The first HybProbe assay carried out was unsuccessful, with no FRET changes observed (Appendix A.16). Amplification did occur, as a product of the expected size was visualised by agarose gel electrophoresis. Lowering of the probe binding temperatures from 60℃ to 56℃, 52℃, and to as low as 45℃, did not improve the assay. A variety of other assay conditions were tested, as listed below.

- **Varying of sensor and probe concentrations**. The sensor and probe concentrations were both increased by 1.5- or 2-fold, or the sensor concentration alone was increased 2- or 3-fold.
- **Change of master mix**. The LightCycler® 480 Genotyping Master (Roche) was used[‡].
- **Addition of additives**. 0.5 μL dH_2O was replaced with 0.5 μL 5x FastStart™ Taq DNA Polymerase GC-RICH Solution (Roche).
- **Change of genotype**. Throughout the assay optimisations, wild-type, heterozygous and homozygous gDNA was used.

[‡]For the 5x LightCycler® 480 Genotyping Master, 2 μL was added in place of 5 μL of the PerfeCTa® qPCR ToughMix®.

None of the changes in conditions improved the assay.

The genomic region surrounding the variant site (within approximately 100 bp either side of the variant) was checked for known common polymorphisms. Additionally, the WES data of individuals from family B was analysed for variants in this region. No known polymorphisms or evidence of variation was discovered, ruling out the possibility that probe binding was inhibited by the presence of an unknown variant.

As fluorescence was not being detected, and amplification problems were ruled out, attention was turned to the integrity of the fluorescent probes. Firstly, the probe concentrations were confirmed to be correct by measurement with the DeNovix spectrophotometer using the ssDNA sample type function. Next, fluorescence of the sensor and anchor probes was measured using the DeNovix DS-11 FX+ fluorometer. The sensor probe (fluorescein) was excited with 470 nm light and the anchor probe (Cy5) with 635 nm light. Multiple measurements, using a range of concentrations, determined that fluorescence of the sensor probe was significantly lower, by a factor of approximately 75-fold, than that of the anchor probe (Appendix A.17).

Because of a lack of sensor probe emission, a PCR was carried out using the sensor probe and the *JSRP1* G150A reverse sequencing primer to test for the presence of the 3' fluorescein modification: polymerisation should occur if there is no 3' modification. The PCR was carried out over three different primer annealing temperatures: 56℃, 58℃ and 60℃. This PCR was carried out alongside a control PCR with both forward and reverse *JSRP1* G150A sequencing primers. The PCR products were visualised by agarose gel electrophoresis (Figure 5.6). No PCR products were observed for the reactions with sensor probe and reverse primer. This did not rule out the possibility that the 3' modification was not present or was damaged, as lack of amplification may have been caused by sub-optimal PCR conditions. The cause of sensor probe problems therefore remained elusive.

The sensor probe was re-synthesised from IDT and the concentration and fluorescence of this new probe was measured on the DeNovix fluorometer as before. There was a ∼3 fold increase in RFU compared to the previously synthesised sensor probe (Appendix A.17), however, the RFU was still significantly lower than the anchor probe.

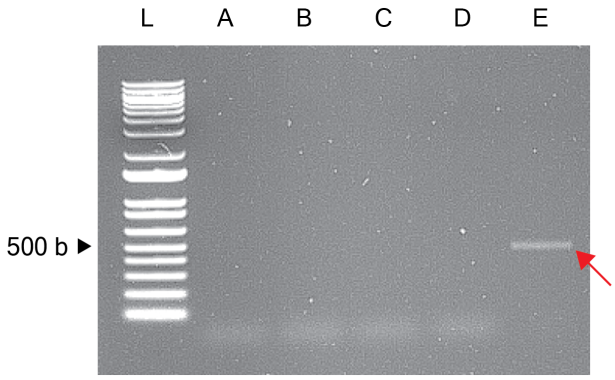

Figure 5.6: Agarose gel electrophoresis of PCR products testing for the presence of a 3' modification of the sensor probe
L is 1 kb+ DNA ladder. A, B and C are PCRs carried out, in the presence of gDNA, with the HybProbe sensor probe and *JSRP1* G150A reverse sequencing primer at annealing temperatures of 56°C, 58°C and 60°C, respectively. D is a negative control carried out with the same conditions as B. E is a positive control carried out with both forward and reverse *JSRP1* G150A sequencing primers. A PCR product of the expected target size, 435 bp, is present in lane E.

5.2.3.4 Error in probe design

At a later time, it was discovered that the sensor probe sequence had been designed incorrectly. The design was intended to complement an 18 nucleotide stretch of DNA that included the G150A variant, however, the G>C variant was incorrectly placed one nucleotide adjacent to the variant position (Figure 5.7).

A) CGCCGTCCCTG**G**GGAG
B) CGCCGTCCCTG**C**GGAG
C) CGCCGTCCCT**C**GGGAG

Figure 5.7: Intended probe design sequence verses actual probe design sequence
A) is the wild-type sequence for the 18 nucleotide stretch. B) is the variant sequence, with the bolded nucleotide being the variant. C) is the actual sequence, with the bolded nucleotide being the incorrectly placed variant.

This would result in a one nucleotide mismatch for the wild-type sequence and a two nucleotide mismatch for the variant sequence. This may explain some of the issues that were encountered with this assay. However, sensor probes are designed to have a one nucleotide mismatch from a wild-type sequence, and so there should be a degree of tolerance for mismatches. As wild-type gDNA was

used when attempting the assay under various conditions, there was only a one nucleotide mismatch with this misdesigned probe.

Another explanation for the lack of fluorescence is that the amplicon or probes may have been forming hair-pins, heterodimers or homodimers. The amplified sequence, and the probes by default, are GC-rich and contain repetitive tri-nucleotide strings GGG and CCC. This had previously made it difficult to design suitable primers to the region surrounding the variant. Even if this was the case, it does not explain the lack of sensor probe emission.

In conclusion, the cause of the assay failure was not determined and an alternative assay was investigated as described in the next section.

5.2.4 Restriction fragment length polymorphism genotyping

Three restriction enzyme cut sites were identified at the variant site: EcoRII, PasI and ScrFI. ScrFI (NEB) was selected to carry out RFLP analysis. Table 5.1 and Figure 5.8 show the expected products after complete digestion by ScrFI for wild-type, homozygous and heterozygous genotypes.

Genotype	**Expected product size/s**
wild-type	81, 25
heterozygous	106, 81, 25
homozygous	106

Table 5.1: Expected fragment sizes upon ScrFI complete digestion
Sizes are given in bp.

Unless otherwise stated, each restriction digest using ScrFI was performed in the following manner: PCR was carried out on gDNA of each sample to be tested for the presence of the variant using HRM primer set 2 (Appendix A.3). The PCR product was visualised by agarose gel electrophoresis for confirmation of successful PCR (a product of 106 bp). One and a half μL 10x CutSmart® (NEB®, Cat No. B7204S) Buffer, 0.4 μL 5,000 units mL^{-1} ScrFI and 13.1 μL PCR product was added to a microcentrifuge tube. The reaction was incubated at 37°C for 60 minutes. Fourteen μL of this reaction was loaded onto a 2.2% TAE

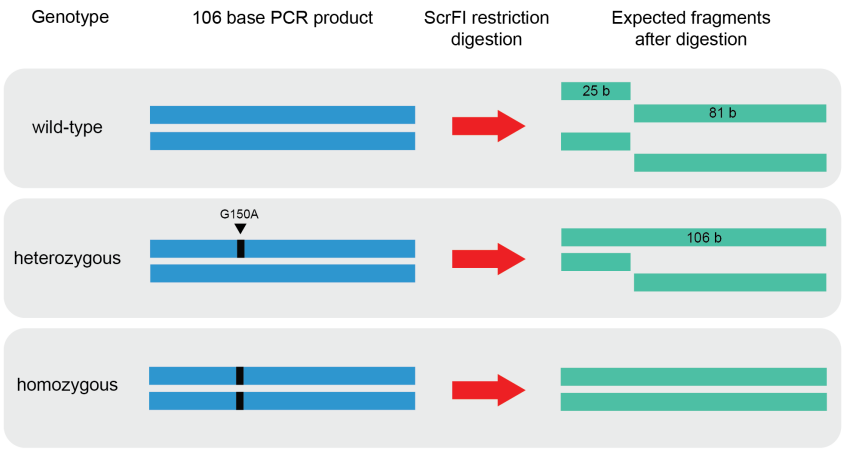

Figure 5.8: Schematic representation of expected fragment sizes upon SrcFI complete digestion

agarose gel containing 0.2 μg mL^{-1} ethidium bromide and electrophoresis was carried out at 100 V for approximately one hour. For each gel, digested genotyping standards were loaded alongside the digested unknowns. The genotyping standards were wild-type 533, homozygous 919 and heterozygous 1772. The gels were visualised under UV light for DNA band visualisation. Genotypes were determined by comparing band patterns with the standards. If no digestion of the 106 bp PCR product was observed, the DNA sample was Sanger sequenced to confirm a homozygous genotype and to disprove the alternative conclusion of a failed restriction digest. A gel image from an initial proof-of-concept restriction digest which was carried out on genotyping standards 533, 919 and 1772 can be seen in Figure 5.9.

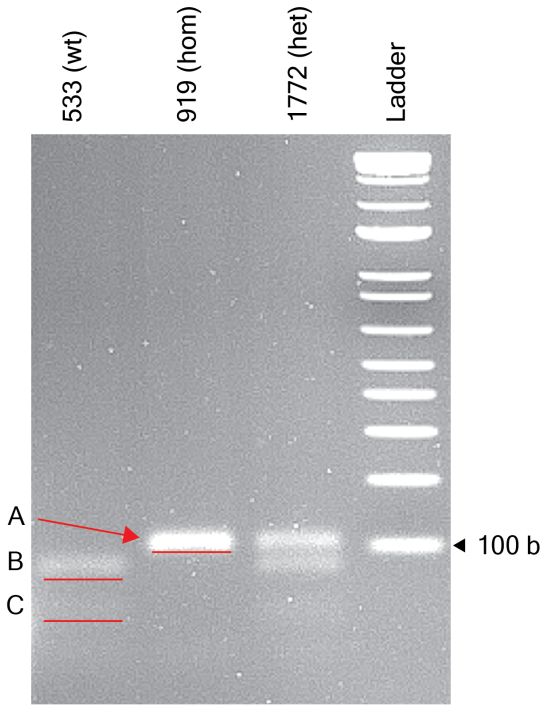

Figure 5.9: Agarose gel electrophoresis of restriction digest products from genotype standards for *JSRP1* G150A RFLP assay
Gel is 2% agarose. Band labelled A is uncut 106 bp PCR product. B is 81 bp digested product. C is 25 bp digested product. 'wt' is wild-type, 'hom' is homozygous, and 'het' is heterozygous. Ladder is 1 kb+ DNA ladder. The red lines aid in identification of the DNA bands.

5.2.5 Statistical analysis

RStudio version 1.2.1335 and ggplot2 version 3.2.1 [136] with R version 3.6.1 was used to perform statistical analyses and create plots (script available at https://github.com/sormond/Masters/scripts/JSRP1_script.R). A linear regression model was used to determine whether *JSRP1* G150A carrier status had a predictor effect on IVCT contracture values. As it has been previously established that *RYR1* T4826I carrier status is a strong predictor of contracture values, an ANOVA test was carried out on the following two linear models (where: c is contracture force in grams, r is *RYR1* T4826I carrier status and j is *JSRP1* G150A carrier status):

$$c \sim r$$

$$c \sim r + j$$

5.3 Results

One hundred and twenty-nine gDNA samples in total were genotyped for *JSRP1* G150A variant. Gel images of PCR and RFLP products for all samples can be found in Appendix A.22, A.23, A.24, A.25, A.26, A.27 and A.29. Tabulated data of all samples and genotypes determined through RFLP analysis are presented in Appendix A.30 and Appendix A.31. The raw data file used for the R script is available in the supplementary material directory of https://github.com/sormond/
Masters (supp_material/ JSRP1_association_data.csv). Of these genotyped samples, 61 were wild-type, 63 were heterozygous, and five were homozygous. Aside from the control homozygous sample 919, three other samples suspected homozygous were Sanger sequenced, which confirmed their genotype (Figure 5.10). Raw data is available

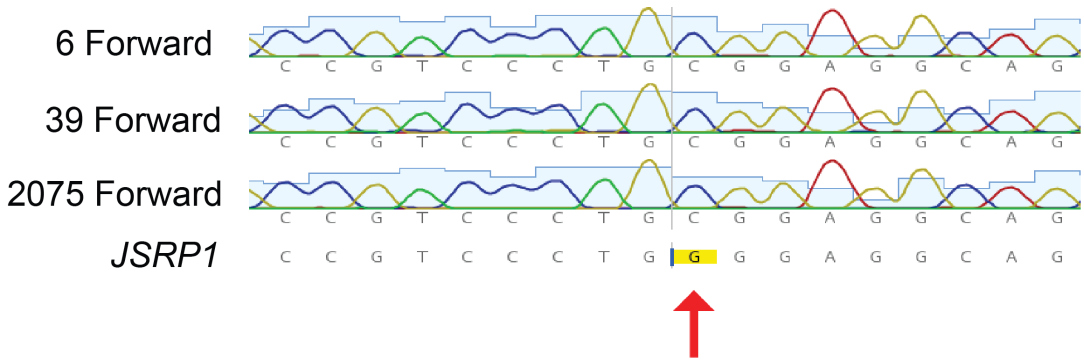

Figure 5.10: Alignment of Sanger sequencing results for individuals 6, 39 and 2075 against the *JSRP1* nucleotide sequence
The *JSRP1* G150A SNV position is indicated by the red arrow. The results indicate that these individuals are homozygous for the variant.

IVCT halothane and caffeine contracture values are plotted against *JSRP1* G150A variant status for all samples in Figure 5.11. Plots for samples that are carriers or not carriers of *RYR1* T4826I variant are given in Figures 5.12 and 5.13, respectively. No significant association between *JSRP1* G150A variant status for either halothane ($F_{1,\ 117} = 0.355$, $P = 0.519$) and caffeine ($F_{1,\ 117} = 0.266$, $P = 0.393$) contracture values was observed [¶].

[¶]Out of 129 genotyped samples, eight were omitted from the statistical calculation because *RYR1* T4826I variant carrier status was unknown for these individuals or because a suspected homozygous genotype was not confirmed by Sanger sequencing.

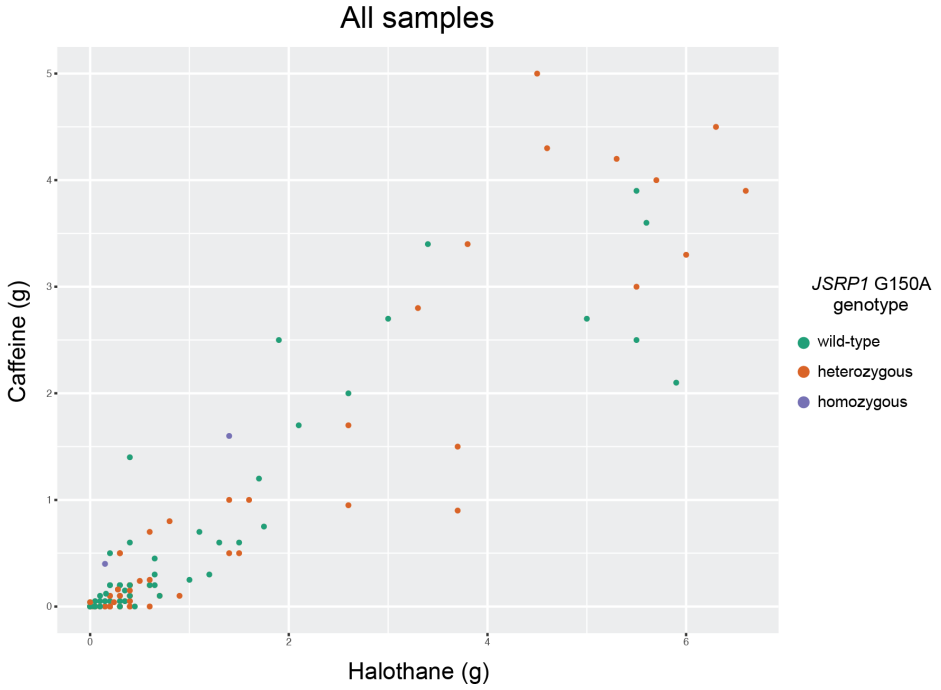

Figure 5.11: IVCT halothane and caffeine contracture force with *JSRP1* G150A genotype for all samples tested

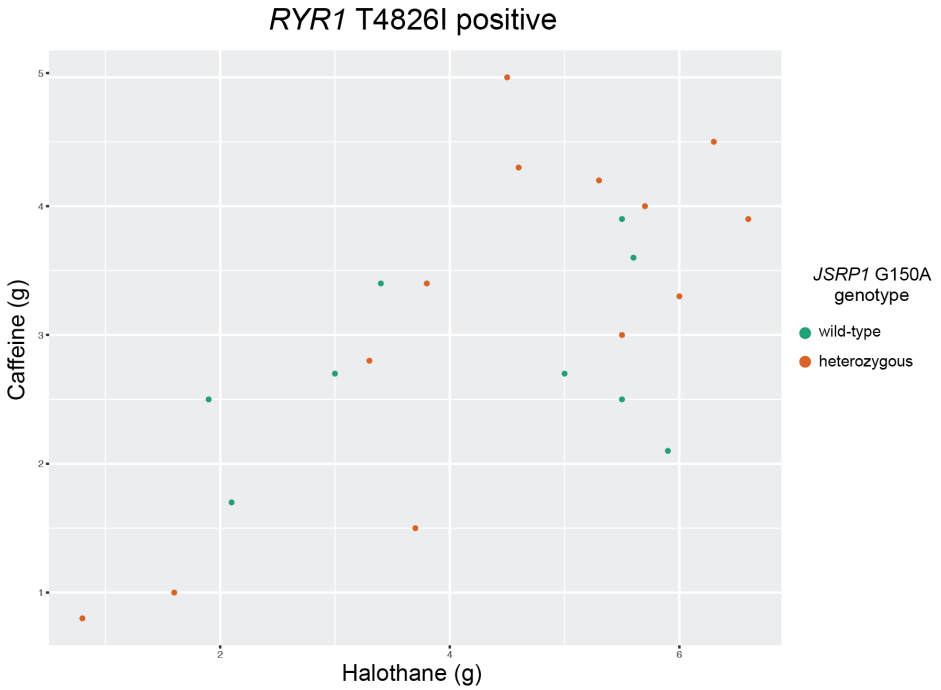

Figure 5.12: IVCT halothane and caffeine contracture force with *JSRP1* G150A genotype for all samples carrying the *RYR1* T4826I variant

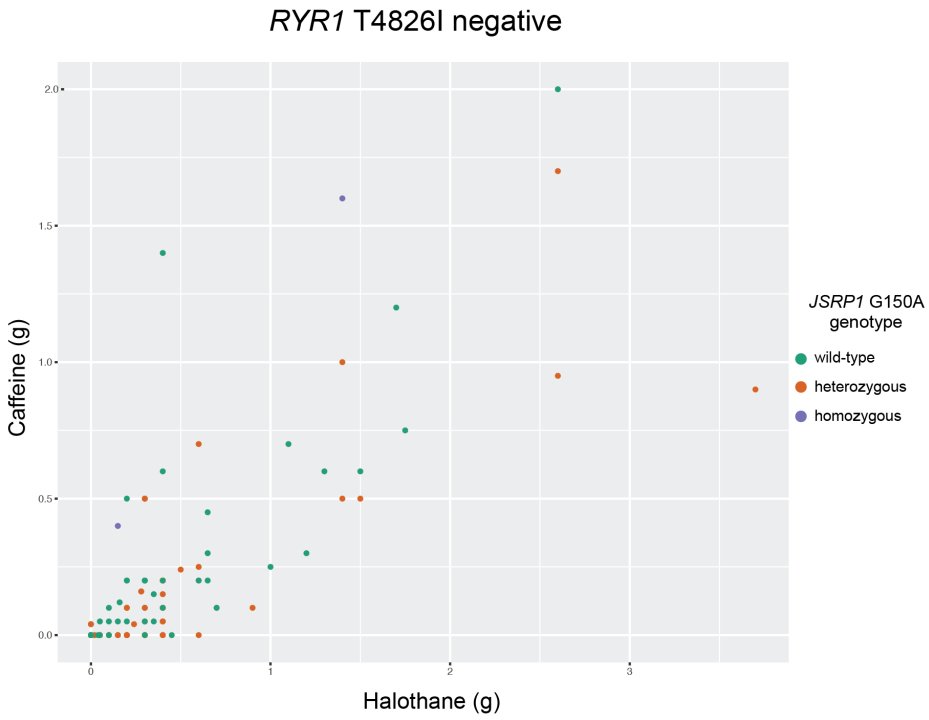

Figure 5.13: IVCT halothane and caffeine contracture force with *JSRP1* G150A genotype for all samples that do not carry the *RYR1* T4826I variant

5.4 Discussion

This study sought to establish whether harbouring the *JSRP1* G150A polymorphism could lead to a reduction in IVCT contracture values. Previous calcium release experiments using human cell lines harbouring the variant showed it to lead to a decrease in the activation sensitivity of DHPR in EC coupling *in vitro*[221], but it was unknown whether this would translate to contracture force measured by IVCT. If this variant was determined to correlate with a decrease in contracture force, it would be the first time a genetic variant which is not suspected to be pathogenic for MH has been demonstrated to alter IVCT values. This is significant because to date, the considerable variation seen in IVCT contracture force has not been successfully attributed to environmental and/or genetic components, aside from suspected and confirmed MH-pathogenic variants. Identifying these other components may help to explain the complexity of MH-susceptibility: For example, why an MH episode is not always triggered in MH-susceptible individuals exposed to triggering volatile anaesthetic agents[42] and the existence of rare heat- and temperature-induced MH episodes[43].

The results of this study were inconsistent with the hypothesis that the *JSRP1* G150A variant is associated with IVCT contracture force for both halothane and caffeine within family B. This does not entirely rule out the possibility that the *JSRP1* G150A polymorphism is a modifying variant however. It is possible that carrying both copies of the polymorphism is necessary to see a significant effect on the results of the IVCT, and with only four confirmed homozygous individuals from family B, it is unlikely that these samples provided the statistical power to see this effect. No *JSRP1* variant homozygous individuals carrying *RYR1* T4826I were identified, and therefore it was not possible to statistically test whether harbouring two copies of the *JSRP1* variant has a modifying effect on *RYR1* T4826I. The results of this experiment cannot be extrapolated to discern whether a genetic variant may modify the effect of certain other MH-related variants (whether that be unknown variants or known MH-pathogenic variants not within family B). Additionally, as the IVCT protocol was designed to sensitively detect MH-susceptibility, the results of the test cannot necessarily be used to determine how a genetic variant may effect the extent of MH-susceptibility within an individual.

It is plausible that the variability seen in IVCT contracture values between individuals is highly dependent on environmental factors, and that genetic elements, aside from MH-causative variants, do not significantly alter these values. Unless any genetic factors contributing to this variability are identified, this will largely remain unknown as it is not feasible to perform more than one IVCT on an individual which would be required to determine the extent to which the environment contributes to the variability.

Chapter 6

Conclusion

It is important that the underlying molecular mechanisms of malignant hyperthermia (MH) be understood so that individuals within these families can be diagnosed by genetic screening as opposed to the *in vitro* contracture test (IVCT), the current invasive diagnostic test employed for MH diagnosis. Additionally, identifying MH-pathogenic variants may improve the current understanding of the inner workings of the myocyte. This could have implications for understanding and treatment of other muscle disorders. Unfortunately, this research did not conclusively uncover variants which seem to strongly associate with MH-susceptibility.

Excellently summarised in Frésard & Montgomery (2018)[223], there are a number of possible reasons that a suspected pathogenic variant may not be detected in WES analysis. These include the following:

1. The pathogenic variant was missed

It may be that an MH-pathogenic variant is present within a final variants of interest list, but that it was discarded as a candidate variant because the protein function of a gene did not seem relevant to MH aetiology. The fact that the cause of MH in almost half of all MH families is unknown suggests that there may be missing pieces to the puzzle in the understanding of MH, and so it is quite plausible that researchers are searching for variants in the wrong genes. It may also have been missed due to a current limited understanding of a gene's function. Even if a protein-coding gene and its product has been characterised, there may be alternative unknown gene transcripts or roles of the protein. Alternatively, a pathogenic variant may fall within one of the number of known genes which are

not captured through WES[224]. Additionally, the identification of mappability issues within the *KNCJ* gene family has suggested that variants within genes that could potentially be of interest to MH research may not be identified.

2. The pathogenic variant does not lie within the exome

One assumption of this study was that the MH-pathogenic variants were present within the exome. Although variants within the exome currently account for the vast majority of known pathogenic variants, it is possible that many pathogenic variants causing disease are being missed because of a focus on the exome. Disease causative non-coding variants are known to impact normal gene regulation[225]. Although it is hard to imagine how a variant which alters a gene's expression could be pathogenic for MH because MH is not a loss-of-function phenotype and known pathogenic MH variants are gain-of-function[56, 55, 42], this avenue should not be ruled out.

3. Monogenic aetiology does not sufficiently explain malignant hyperthermia

A recurring issue in MH genetics is the presence of discordance[226, 130] within families with variants which are believed to be pathogenic for MH. This could be due to misdiagnosis by IVCT[62], but another explanation could be the presence of currently unknown disease modifying variants. This could be preventing the identification of MH-pathogenic variants. A disease modifying variant may alter the expression of a gene harbouring a pathogenic variant, and thus could affect the disease phenotype. For instance, the phenotype of an MH-pathogenic variant could be masked by a second variant, leading to an MHN diagnosis in an individual carrying the MH-pathogenic variant. This would prevent the identification of the pathogenic variant, as it is currently assumed that MHN individuals will not carry an unknown MH-pathogenic variant. To account for this possibility, some degree of lenience should be integrated into the gtFilter programme used in the bioinformatic pipeline. Currently, gtFilter only accounts for the possibility of the presence of individuals incorrectly diagnosed as MHS. Importantly however, larger cohorts of individuals for both bioinformatic analyses to identify VOI and segregation analyses would be necessary. It would be extremely difficult to identify modifying variants, especially if a modifier was harboured by only one or a handful of individuals within a family, however there is potential that greater future understanding of molecular processes governing susceptibility to MH will

aid in the discovery of modifying genetic elements.

There is one variant, *JSRP1* G150A, which, at present, is the only common polymorphism postulated to be disease-modifying for MH. A genotype-phenotype association study was carried out on the variant, previously associated with reduced strength of excitation-contraction coupling upon a stimulation event[221], which sought to establish whether the variant could have an affect on IVCT contracture responses. The polymorphism did not appear to be associated with altered IVCT contracture values to either caffeine and halothane in the family investigated.

Regarding the bioinformatic aspect of the project, this research has highlighted the importance of taking both the disorder under study and the sequencing data at hand into consideration when establishing a suitable bioinformatic pipeline. In this case, a strict variant filtering system could not be used because the diagnostic method for determining phenotype was imperfect[62]. Additionally, the establishment of this bioinformatic pipeline required an understanding of the quality of the whole exome sequencing (WES) data and the impact that this would have on the ability to capture VOI. Somewhat uniquely, this study involved the use of WES data from a range of technologies over a number of years. It was quickly established that this had implications on the identification and filtering of variants, as both low WES data quality and lack of agreement between WES kit target regions meant that potential VOI were not identifiable in some samples. It was important that the bioinformatic pipeline retain any VOI, even if they were not identified, due to lack of coverage, in most of the samples. One good example of the effectiveness of the established bioinformatic pipeline was the identification of a *WASHC2C* variant in family B, in which only two WES samples were genotyped due to a lack of coverage at the variant region for other family B samples; a potential association between *WASHC2C* variation and MH-susceptibility was recognised because of this, as variants within this gene were also identified in families A and C.

6.1 Future directions

6.1.1 Segregation analysis of identified variants

Confirmation of the presence of a variant by Sanger sequencing, followed by segregation analysis using high resolution melting analysis, should be performed for the other VOI identified in this project. For the identified copy number variants (CNV), performing CNV analysis using CNVkit [126] as described on other WES data from the families would be necessary to ascertain whether a CNV is segregating with MH-susceptibility. If segregation amongst the WES samples was established, it would be useful to carry out a more comprehensive literature search to see whether this CNV has been previously reported. Then, confirmation of the presence of the CNV using gDNA from the WES-sampled individuals could be carried out by quantitative-PCR.

6.1.2 Carrying out further sequencing

Increasing the number of sequenced samples will increase the probability of identifying segregating variants of interest. If the pathogenic variant was missed because of a lack of knowledge of protein function and/or MH pathomechanisms, undertaking further sequencing on the families, and on other MH families, will increase the likelihood of identification of common genes and variants. Additionally, it could be beneficial to carry out whole genome sequencing, as this would provide an opportunity to identify segregating CNVs and non-coding variants, however this would be significantly more expensive and would require modification of the current bioinformatic pipeline. Resequencing should be undertaken on family D and family F due to the very low WES coverage.

6.1.3 Future reanalysis of WES data

The constant improvement in sequencing technologies, and in the collective understanding of the human genome and human genetic disease, will translate to improved sequencing data, bioinformatic tools, and bioinformatic resources. This makes it imperative that reanalysis of WES data, as suggested by Nambot et al. (2018) [129], be undertaken in the future. Reanalysis would involve analysing the WES data using the bioinformatic pipeline with the most recently released versions and databases, including RefSeq [102] (gene transcripts), the Gene

Ontology[116,117] (known gene functions), dbSNP[119] (known variants), the Human Protein Atlas[121] (tissue gene expression), as well as other categorical and functional annotations applied to variants by VarAFT[123]. Updated versions of software, such as the Genome Analysis Toolkit[94], used for processing human genomic sequencing data should also be used to ensure WES data processing follows future best practices. Additionally, it would be beneficial to apply updated allele frequencies to the variants, especially if large-scale genome sequencing is undertaken on Māori or Pasifika populations. This has the potential to greatly decrease the number of polymorphisms that appear rare within Māori or other under-represented ethnicities due to the current European bias present within allele frequency databases. To account for the possibility that a disease modifying variant may be hindering identification of an MH-pathogenic variant by masking an MHS phenotype, the gtFilter programme should be updated to include a leniency for MHN individuals.

6.1.4 Functional analysis of segregating variants

If a variant is identified and discovered to be segregating with MH-susceptibility within a family, functional analysis of the effected protein should be undertaken. Functional characterisation is a requirement for acceptance of a diagnostic MH variant according to European Malignant Hyperthermia Group guidelines[45]. How this functional analysis could be carried out would be dependent on the affected protein. Currently, *RYR1* variants are functionally characterised *in vitro* using human HEK293T cell lines by transfection with *RYR1* cDNA harbouring the VOI[55,56]. Ca^{2+} release is measured upon treatment with RyR1 agonists. However, these cells do not accurately mimic *in vivo* myocytes as they do not have a sarcoplasmic reticulum and do not express skeletal muscle DHPR and other proteins involved in the excitation-contraction coupling machinery. This is a current major limitation in the functional characterisation of non-*RYR1* variants. However, if a segregating variant was identified in a gene encoding a protein which is believed to interact and functionally alter RyR1's properties, Ca^{2+} release assays could be carried out on a human cell line co-transfected with cDNA of *RYR1* and the gene of interest. Additionally, a mouse knock-out of the gene in question, followed by knock-in of the variant gene, could help to dissect protein function and and inform *in vitro* functional studies.

6.2 Concluding remarks

Although this study has not uncovered a variant which strongly associates with MH in any of the six families under study, the work described in this thesis has provided a sound platform for future studies. A bioinformatic pipeline capable of identifying VOI and filtering them in a phenotype-specific manner, whilst also taking into account potential inaccurate diagnoses of the phenotype, has been established. This pipeline has been demonstrated to be robust and can be used to study other uncharacterised MH families. Additionally, segregation analysis has not yet been carried out for a number of identified VOI, providing further research avenues.

Although there have been a number of VOI that have been ruled out as being pathogenic for MH-susceptibility, it is important to bear in mind that there have been occurrences of MHN individuals harbouring familial variants that are known to be pathogenic for MH[227]. Additional gDNA and phenotypic data will be obtained from the families in future, providing the opportunity to perform larger studies on segregation for VOI, and to carry out further WES or WGS for bioinformatic analyses. In the meantime, the variants of interest identified in this study should remain as such, simply variants of interest.

Bibliography

[1] D. Richfield. Medical gallery of david richfield. https://en.wikiversity.org/wiki/WikiJournal_of_Medicine/Medical_gallery_of_David_Richfield_2014, 2014. ISSN ISSN 2002-4436. Accessed on 2020-08-02.

[2] Blausen.com staff. Medical gallery of blausen medical. https://en.wikiversity.org/wiki/WikiJournal_of_Medicine/Medical_gallery_of_Blausen_Medical_2014, 2014. ISSN 2002-4436. Accessed on 2019-05-23.

[3] E. Rios and G. Brum. Involvement of dihydropyridine receptors in excitation contraction coupling in skeletal-muscle. *Nature*, 325(6106):717–720, FEB 19 1987. ISSN 0028-0836. doi: {10.1038/325717a0}.

[4] C. Proenza, J. O'Brien, J. Nakai, S. Mukherjee, P. D. Allen, and K. G. Beam. Identification of a region of RyR1 that participates in allosteric coupling with the alpha(1s) (Ca(V)1.1) II-III loop. *Journal Of Biological Chemistry*, 277(8):6530–6535, FEB 22 2002. ISSN 0021-9258. doi: {10.1074/jbc.M106471200}.

[5] B. R. Nelson, F. Wu, Y. Liu, D. M. Anderson, J. McAnally, W. Lin, S. C. Cannon, et al. Skeletal muscle-specific T-tubule protein STAC3 mediates voltage-induced Ca^{2+} release and contractility. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29):11881–11886, JUL 16 2013. ISSN 0027-8424. doi: {10.1073/pnas.1310571110}.

[6] A. A. Anderson, X. Altafaj, Z. Zheng, Z. Wang, O. Delbono, M. Ronjat, S. Treves, et al. The junctional SR protein JP-45 affects the functional expression of the voltage-dependent Ca^{2+} channel Ca(v)1.1. *Journal of Cell Science*, 119(10):2145–2155, MAY 15 2006. ISSN 0021-9533. doi: {10.1242/jcs.02935}.

[7] M. Fill and J. A. Copello. Ryanodine receptor calcium release channels. *Physiological Reviews*, 82(4):893–922, OCT 2002. ISSN 0031-9333. doi: {10.1152/physrev.00013.2002}.

[8] Z. Yan, X. Bai, C. Yan, J. Wu, Z. Li, T. Xie, W. Peng, et al. Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution. *Nature*, 517 (7532):50+, JAN 1 2015. ISSN 0028-0836. doi: {10.1038/nature14063}.

[9] G. G. Rodney, B. Y. Williams, G. M. Strasburg, K. Beckingham, and S. L. Hamilton. Regulation of RYR1 activity by Ca^{2+} and calmodulin. *Biochemistry*, 39(26):7807–7812, JUL 4 2000. ISSN 0006-2960. doi: {10.1021/bi0005660}.

[10] L. W. Xiong, J. Z. Zhang, R. He, and S. L. Hamilton. A Ca^{2+}-binding domain in RyR1 that interacts with the calmodulin binding site and modulates channel activity. *Biophysical Journal*, 90(1):173–182, JAN 2006. ISSN 0006-3495. doi: {10.1529/biophysj.105.066092}.

[11] R. T. Rebbeck, F. R. Nitu, D. Rohde, P. Most, D. M. Bers, D. D. Thomas, and R. L. Cornea. S100A1 protein does not compete with calmodulin for ryanodine receptor binding but structurally alters the ryanodine receptor center dot calmodulin complex. *Journal Of Biological Chemistry*, 291(30): 15896–15907, JUL 22 2016. doi: {10.1074/jbc.M115.713107}.

[12] W. Guo and K. P. Campbell. Association of triadin with the ryanodine receptor and calsequestrin in the lumen of the sarcoplasmic-reticulum. *Journal Of Biological Chemistry*, 270(16):9027–9030, APR 21 1995. ISSN 0021-9258. doi: {10.1074/jbc.270.16.9027}.

[13] N. A. Beard, M. M. Sakowska, A. F. Dulhunty, and D. R. Laver. Calsequestrin is an inhibitor of skeletal muscle ryanodine receptor calcium release channels. *Biophysical Journal*, 82(1):310–320, JAN 2002. ISSN 0006-3495. doi: {10.1016/S0006-3495(02)75396-4}.

[14] N. A. Beard and A. F. Dulhunty. C-terminal residues of skeletal muscle calsequestrin are essential for calcium binding and for skeletal ryanodine receptor inhibition. *Skeletal Muscle*, 5, FEB 22 2015. ISSN 2044-5040. doi: {10.1186/s13395-015-0029-7}.

[15] L. Wei, E. M. Gallant, A. F. Dulhunty, and N. A. Beard. Junctin and triadin each activate skeletal ryanodine receptors but junctin alone mediates functional interactions with calsequestrin. *International Journal of Biochemistry & Cell Biology*, 41(11):2214–2224, NOV 2009. ISSN 1357-2725. doi: {10.1016/j.biocel.2009.04.017}.

[16] S. Boncompagni, M. Thomas, J. R. Lopez, P. D. Allen, Q. Yuan, E. G. Kranias, C. Franzini-Armstrong, et al. Triadin/junctin double null mouse reveals a differential role for triadin and junctin in anchoring CASQ to the JSR and regulating Ca^{2+} homeostasis. *PLOS One*, 7(7), JUL 2 2012. ISSN 1932-6203. doi: {10.1371/journal.pone.0039962}.

[17] B. Mosca, J. Eckhardt, L. Bergamelli, S. Treves, R. Bongianino, M. De Negri, S. G. Priori, et al. Role of the JP45-calsequestrin complex on calcium entry in slow twitch skeletal muscles. *Journal Of Biological Chemistry*, 291 (28):14555–14565, JUL 8 2016. doi: {10.1074/jbc.M115.709071}.

[18] T. Wang, L. Zhang, C. Shi, R. Wei, and C. Yin. Interaction of the Homer1 EVH1 domain and skeletal muscle ryanodine receptor. *Biochemical and Biophysical Research Communications*, 514(3):720–725, JUN 30 2019. ISSN 0006-291X. doi: {10.1016/j.bbrc.2019.04.199}.

[19] W. Feng, J. C. Tu, T. Z. Yang, P. S. Vernon, P. D. Allen, P. F. Worley, and I. N. Pessah. Homer regulates gain of ryanodine receptor type 1 channel complex. *Journal Of Biological Chemistry*, 277(47):44722–44730, NOV 22 2002. ISSN 0021-9258. doi: {10.1074/jbc.M207675200}.

[20] W. Feng, J. Tu, P. Pouliquin, E. Cabrales, X. Shen, A. Dulhunty, P. F. Worley, et al. Dynamic regulation of ryanodine receptor type 1 (RyR1) channel activity by Homer 1. *Cell Calcium*, 43(3):307–314, MAR 2008. ISSN 0143-4160. doi: {10.1016/j.ceca.2007.06.001}.

[21] G. Avila, E. H. Lee, C. F. Perez, P. D. Allen, and R. T. Dirksen. FKBP12 binding to RyR1 modulates excitation-contraction coupling in mouse skeletal myotubes. *Journal Of Biological Chemistry*, 278(25):22600–22608, JUN 20 2003. ISSN 0021-9258. doi: {10.1074/jbc.M205866200}.

[22] M. Gaburjakova, J. Gaburjakova, S. Reiken, F. Huang, S. O. Marx, N. Rosemblit, and A. R. Marks. FKBP12 binding modulates ryanodine

receptor channel gating. *Journal Of Biological Chemistry*, 276(20):16931–16935, MAY 18 2001. ISSN 0021-9258. doi: {10.1074/jbc.M100856200}.

[23] S. Reiken, A. Lacampagne, H. Zhou, A. Kherani, S. E. Lehnart, C. Ward, F. Huang, et al. PKA phosphorylation activates the calcium release channel (ryanodine receptor) in skeletal muscle: defective regulation in heart failure. *Journal of Cell Biology*, 160(6):919–928, MAR 17 2003. ISSN 0021-9525. doi: {10.1083/jcb.200211012}.

[24] A. M. Bellinger, S. Reiken, M. Dura, P. W. Murphy, S. Deng, D. W. Landry, D. Nieman, et al. Remodeling of ryanodine receptor complex causes "leaky" channels: A molecular mechanism for decreased exercise capacity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(6):2198–2202, FEB 12 2008. ISSN 0027-8424. doi: {10.1073/pnas.0711074105}.

[25] P. Aracena, W. T. Tang, S. L. Hamilton, and C. Hidalgo. Effects of S-glutathionylation and S-nitrosylation on calmodulin binding to triads and FKBP12 binding to type 1 calcium release channels. *Antioxidants & Redox Signaling*, 7(7-8):870–881, JUL-AUG 2005. ISSN 1523-0864. doi: {10.1089/ars.2005.7.870}.

[26] P. Aracena-Parks, S. A. Goonasekera, C. P. Gilman, R. T. Dirksen, C. Hidalgo, and S. L. Hamilton. Identification of cysteines involved in S-nitrosylation, S-glutathionylation, and oxidation to disulfides in ryanodine receptor type 1. *Journal Of Biological Chemistry*, 281(52):40354–40368, DEC 29 2006. doi: {10.1074/jbc.M600876200}.

[27] Q. Sun, B. Wang, M. Miyagi, D. T. Hess, and J. S. Stamler. Oxygen-coupled redox regulation of the skeletal muscle ryanodine receptor/Ca^{2+} release channel (RyR1) sites and nature of oxidative modification. *Journal Of Biological Chemistry*, 288(32):22961–22971, AUG 9 2013. doi: {10.1074/jbc.M113.480228}.

[28] T. Yamada, O. Fedotovskaya, A. J. Cheng, A. S. Cornachione, F. C. Minozzo, C. Aulin, C. Friden, et al. Nitrosative modifications of the Ca^{2+} release complex and actin underlie arthritis-induced muscle weakness. *Annals of the Rheumatic Diseases*, 74(10):1907–1914, OCT 2015. ISSN 0003-4967. doi: {10.1136/annrheumdis-2013-205007}.

[29] J. M. Dias and P. D. Vogel. Effects of small molecule modulators on ATP binding to skeletal ryanodine receptor. *Protein Journal*, 28(5):240–246, JUN 2009. ISSN 1572-3887. doi: {10.1007/s10930-009-9189-9}.

[30] D. R. Laver. Ca^{2+} stores regulate ryanodine receptor Ca^{2+} release channels via luminal and cytosolic Ca^{2+} sites. *Clinical and Experimental Pharmacology and Physiology*, 34(9):889–896, SEP 2007. ISSN 1440-1681. doi: {10.1111/j.1440-1681.2007.04708.x}.

[31] J. A. Copello, S. Barg, A. Sonnleitner, M. Porta, P. Diaz-Sylvester, M. Fill, H. Schindler, and S. Fleischer. Differential activation by Ca^{2+}, ATP and caffeine of cardiac and skeletal muscle ryanodine receptors after block by Mg^{2+}. *Journal of Membrane Biology*, 187(1):51–64, MAY 1 2002. ISSN 0022-2631. doi: {10.1007/s00232-001-0150-x}.

[32] B. R. MacIntosh, R. J. Holash, and J. Renaud. Skeletal muscle fatigue - regulation of excitation-contraction coupling to avoid metabolic catastrophe. *Journal of Cell Science*, 125(9):2105–2114, MAY 1 2012. ISSN 0021-9533. doi: {10.1242/jcs.093674}.

[33] S. Burling. *Contributions of calstabins acting on RYR1 in malignant hyperthermia susceptibility: a thesis presented in partial fulfilment of the requirements for the degree of Bachelor of Science with Honours in Biochemistry at Massey University, Manawatu, New Zealand.* 2019.

[34] D. H. MacLennan, W. J. Rice, and N. M. Green. The mechanism of Ca^{2+} transport by sarco(endo)plasmic reticulum Ca^{2+}-ATPases. *Journal Of Biological Chemistry*, 272(46):28815–28818, NOV 14 1997. ISSN 0021-9258. doi: {10.1074/jbc.272.46.28815}.

[35] I. C. Smith, E. Bombardier, C. Vigna, and A. R. Tupling. ATP consumption by sarcoplasmic reticulum Ca^{2+} pumps accounts for 40-50% of resting metabolic rate in mouse fast and slow twitch skeletal muscle. *PLOS One*, 8(7), JUL 1 2013. ISSN 1932-6203. doi: {10.1371/journal.pone.0068924}.

[36] Y. Kimura, K. Kurzydlowski, M. Tada, and D. H. MacLennan. Phospholamban regulates the Ca^{2+}-ATPase through intramembrane interactions. *Journal Of Biological Chemistry*, 271(36):21726–21731, SEP 6 1996. ISSN 0021-9258. doi: {10.1074/jbc.271.36.21726}.

[37] A. Odermatt, S. Becker, V. K. Khanna, K. Kurzydlowski, E. Leisner, D. Pette, and D. H. MacLennan. Sarcolipin regulates the activity of SERCA1, the fast-twitch skeletal muscle sarcoplasmic reticulum Ca^{2+}-ATPase. *Journal Of Biological Chemistry*, 273(20):12360–12369, MAY 15 1998. ISSN 0021-9258. doi: {10.1074/jbc.273.20.12360}.

[38] J. Soboloff, B. S. Rothberg, M. Madesh, and D. L. Gill. STIM proteins: dynamic calcium signal transducers. *Nature Reviews Molecular Cell Biology*, 13(9):549–565, SEP 2012. ISSN 1471-0072. doi: {10.1038/nrm3414}.

[39] D. E. Clapham. Calcium signaling. *Cell*, 131(6):1047–1058, DEC 14 2007. ISSN 0092-8674. doi: {10.1016/j.cell.2007.11.028}.

[40] B. Nilius and A. Szallasi. Transient receptor potential channels as drug targets: from the science of basic research to the art of medicine. *Pharmacological Reviews*, 66(3):676–814, JUL 2014. ISSN 0031-6997. doi: {10.1124/pr.113.008268}.

[41] C. Vandebrouck, D. Martin, M. C. V. Schoor, H. Debaix, and P. Gailly. Involvement of TRPC in the abnormal calcium influx observed in dystrophic (mdx) mouse skeletal muscle fibers. *Journal of Cell Biology*, 158(6):1089–1096, SEP 16 2002. ISSN 0021-9525. doi: {10.1083/jcb.200203091}.

[42] H. Rosenberg, N. Pollock, A. Schiemann, T. Bulger, and K. Stowell. Malignant hyperthermia: a review. *Orphanet Journal Of Rare Diseases*, 10, AUG 4 2015.

[43] K. M. Stowell. Malignant hyperthermia: a pharmacogenetic disorder. *Pharmacogenomics*, 9(11):1657–1672, NOV 2008. ISSN 1462-2416. doi: {10.2217/14622416.9.11.1657}.

[44] T. V. Mccarthy, J. M. S. Healy, J. Heffron, M. Lehane, T. Deufel, F. Lehmannhorn, M. Farrall, et al. Localization of the malignant hyperthermia susceptibility locus to human-chromosome 19q12-13.2. *Nature*, 343 (6258):562–564, FEB 8 1990. ISSN 0028-0836. doi: {10.1038/343562a0}.

[45] P. M. Hopkins, H. Rueffert, M. M. Snoeck, T. Girard, K. P. E. Glahn, F. R. Ellis, C. R. Mueller, et al. European Malignant Hyperthermia Group guidelines for investigation of malignant hyperthermia susceptibility. *British*

Journal of Anaesthesia, 115(4):531–539, OCT 2015. ISSN 0007-0912. doi: {10.1093/bja/aev225}.

[46] R. C. Levitt, N. Nouri, A. E. Jedlicka, V. A. Mckusick, A. R. Marks, J. G. Shutack, J. E. Fletcher, et al. Evidence for genetic-heterogeneity in malignant hyperthermia susceptibility. *Genomics*, 11(3):543–547, NOV 1991. ISSN 0888-7543. doi: {10.1016/0888-7543(91)90061-I}.

[47] D. E. Iles, F. Lehmannhorn, S. W. Scherer, L. C. Tsui, D. O. Weghuis, R. F. Suijkerbuijk, L. Heytens, et al. Localization of the gene encoding the alpha(2)/delta-subunits of the L-type voltage-dependent calcium-channel to chromosome 7q and analysis of the segregation of flanking markers in malignant hyperthermia susceptible families. *Human Molecular Genetics*, 3(6):969–975, JUN 1994. ISSN 0964-6906. doi: {10.1093/hmg/3.6.969}.

[48] R. Sudbrak, V. Procaccio, M. Klausnitzer, J. L. Curran, K. Monsieurs, C. Van Broeckhoven, R. Ellis, et al. Mapping of a further malignant hyperthermia susceptibility locus to chromosome 3q13.1. *American Journal of Human Genetics*, 56(3):684–691, MAR 1995. ISSN 0002-9297.

[49] R. L. Robinson, N. Monnier, W. Wolz, M. Jung, A. Reis, G. Nuernberg, J. L. Curran, et al. A genome wide search for susceptibility loci in three European malignant hyperthermia pedigrees. *Human Molecular Genetics*, 6(6):953–961, JUN 1997. ISSN 0964-6906. doi: {10.1093/hmg/6.6.953}.

[50] J. Nakai, T. Tanabe, T. Konno, B. Adams, and K. G. Beam. Localization in the II-III loop of the dihydropyridine receptor of a sequence critical for excitation-contraction coupling. *Journal Of Biological Chemistry*, 273(39): 24983–24986, SEP 25 1998. ISSN 0021-9258. doi: {10.1074/jbc.273.39.24983}.

[51] J. H. Kim, G. P. Jarvik, B. L. Browning, R. Rajagopalan, A. S. Gordon, M. J. Rieder, P. D. Robertson, et al. Exome sequencing reveals novel rare variants in the ryanodine receptor and calcium channel genes in malignant hyperthermia families. *Anesthesiology*, 119(5):1054–1065, NOV 2013. ISSN 0003-3022. doi: {10.1097/ALN.0b013e3182a8a998}.

[52] D. Fiszer, M. Shaw, N. A. Fisher, I. M. Carr, P. K. Gupta, E. J. Watkins, D. R. de Sa, et al. Next-generation sequencing of RYR1 and CACNA1S

in malignant hyperthermia and exertional heat illness. *Anesthesiology*, 122(5):1033–1046, MAY 2015. ISSN 0003-3022. doi: {10.1097/ALN.0000000000000610}.

[53] C. F. Perez, J. M. Eltit, J. R. Lopez, D. Bodnar, A. F. Dulhunty, S. Aditya, and M. G. Casarotto. Functional and structural characterization of a novel malignant hyperthermia-susceptible susceptible variant of DHPR-beta(1a) subunit (CACNB1). *American Journal of Physiology- Cell Physiology*, 314 (3):C323–C333, MAR 2018. ISSN 0363-6143. doi: {10.1152/ajpcell.00187.2017}.

[54] E. J. Horstick, J. W. Linsley, J. J. Dowling, M. A. Hauser, K. K. McDonald, A. Ashley-Koch, L. Saint-Amant, et al. Stac3 is a component of the excitation-contraction coupling machinery and mutated in Native American myopathy. *Nature Communications*, 4, JUN 2013. ISSN 2041-1723. doi: {10.1038/ncomms2952}.

[55] A. H. Schiemann, N. Paul, R. Parker, N. Pollock, T. F. Bulger, and K. M. Stowell. Functional characterization of 2 known ryanodine receptor mutations causing malignant hyperthermia. *Anesthesia and Analgesia*, 118(2):375–380, FEB 2014. ISSN 0003-2999. doi: {10.1213/ANE.0b013e3182a273ea}.

[56] C. Roesl, K. Sato, A. Schiemann, N. Pollock, and K. M. Stowell. Functional characterisation of the R2452W ryanodine receptor variant associated with malignant hyperthermia susceptibility. *Cell Calcium*, 56(3):195–201, SEP 2014. ISSN 0143-4160. doi: {10.1016/j.ceca.2014.07.004}.

[57] O. Mungunsukh, P. Deuster, S. Muldoon, F. O'Connor, and N. Sambuughin. Estimating prevalence of malignant hyperthermia susceptibility through population genomics data. *British Journal of Anaesthesia*, 123(3): E461–E463, SEP 2019. ISSN 0007-0912. doi: {10.1016/j.bja.2019.06.007}.

[58] B. A. Britt and W. Kalow. Malignant hyperthermia - a statistical review. *Canadian Anaesthetists Society Journal*, 17(4):293+, 1970. ISSN 0008-2856. doi: {10.1007/BF03004694}.

[59] H. Ording. IIncidence of malignant hyperthermia in Denmark. *Anesthesia and Analgesia*, 64(7):700–704, 1985. ISSN 0003-2999.

[60] F. R. Ellis. A protocol for the investigation of malignant hyperpyrexia (MH) susceptibility. *British Journal of Anaesthesia*, 56(11):1267–1269, 1984. ISSN 0007-0912.

[61] N. Pollock, E. E. Langton, K. M. Stowell, and T. F. Bulger. Safety of exposure of malignant hyperthermia non-susceptible patients and their relatives to anaesthetic triggering agents. *Anaesthesia and Intensive Care*, 39(5):887–894, SEP 2011. ISSN 0310-057X. doi: {10.1177/0310057X1103900514}.

[62] H. Ording, V. Brancadoro, S. Cozzolino, F. R. Ellis, V. Glauber, E. F. Gonano, P. J. Halsall, et al. In vitro contracture test for diagnosis of malignant hyperthermia following the protocol of the European MH Group: Results of testing patients surviving fulminant MH and unrelated low-risk subjects. *Acta Anaesthesiologica Scandinavica*, 41(8):955–966, SEP 1997. ISSN 0001-5172. doi: {10.1111/j.1399-6576.1997.tb04820.x}.

[63] R. M. Quinlivan, C. R. Muller, M. Davis, N. G. Laing, G. A. Evans, J. Dwyer, J. Dove, et al. Central core disease: clinical, pathological, and genetic features. *Archives of Disease in Childhood*, 88(12):1051–1055, DEC 2003. ISSN 0003-9888. doi: {10.1136/adc.88.12.1051}.

[64] S. Wu, M. C. A. Ibarra, M. C. V. Malicdan, K. Murayama, Y. Ichihara, H. Kikuchi, I. Nonaka, et al. Central core disease is due to RYR1 mutations in more than 90
patients. *Brain*, 129(6):1470–1480, 04 2006. ISSN 0006-8950. doi: 10.1093/brain/awl077. URL `https://doi.org/10.1093/brain/awl077`.

[65] P. J. Lynch, J. F. Tong, M. Lehane, A. Mallet, L. Giblin, J. A. Heffron, P. Vaughan, et al. A mutation in the transmembrane/luminal domain of the ryanodine receptor is associated with abnormal Ca^{2+} release channel function and severe central core disease. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7):4164–4169, MAR 30 1999. ISSN 0027-8424. doi: {10.1073/pnas.96.7.4164}.

[66] N. Tilgen, F. Zorzato, B. Halliger-Keller, F. Muntoni, C. Sewry, L. M. Palmucci, C. Schneider, et al. Identification of four novel mutations in the C-terminal membrane spanning domain of the ryanodine receptor 1:

association with central core disease and alteration of calcium homeostasis. *Human Molecular Genetics*, 10(25):2879–2887, DEC 1 2001. ISSN 0964-6906. doi: {10.1093/hmg/10.25.2879}.

[67] J. J. Dowling, S. Lillis, K. Amburgey, H. Zhou, S. Al-Sarraj, S. J. A. Buk, E. Wraige, et al. King-Denborough syndrome with and without mutations in the skeletal muscle ryanodine receptor (RYR1) gene. *Neuromuscular Disorders*, 21(6):420–427, JUN 2011. ISSN 0960-8966. doi: {10.1016/j.nmd.2011.03.006}.

[68] S. Guis, D. Figarella-Branger, N. Monnier, D. Bendahan, G. Kozak-Ribbens, J. P. Mattei, J. Lunardi, et al. Multiminicore disease in a family susceptible to malignant hyperthermia - Histology, in vitro contracture tests, and genetic characterization. *Archives of Neurology*, 61(1):106–113, JAN 2004. ISSN 0003-9942. doi: {10.1001/archneur.61.1.106}.

[69] J. M. Wilmshurst, S. Lillis, H. Zhou, K. Pillay, H. Henderson, W. Kress, C. R. Mueller, et al. RYR1 mutations are a common cause of congenital myopathies with central nuclei. *Annals of Neurology*, 68(5):717–726, NOV 2010. ISSN 0364-5134. doi: {10.1002/ana.22119}.

[70] A. Gottschalk, T. Heiman-Patterson, R. deQuevedo, and P. D. Quinn. General anesthesia for a patient with centronuclear (myotubular) myopathy. *Anesthesiology*, 89(4):1018–1020, OCT 1998. ISSN 0003-3022. doi: {10.1097/00000542-199810000-00026}.

[71] J. A. Morgan-Hughes. Tubular aggregates in skeletal muscle: their functional significance and mechanisms of pathogenesis. *Current Opinion in Neurology*, 11(5):439–442, OCT 1998. ISSN 1350-7540. doi: {10.1097/00019052-199810000-00005}.

[72] J. Lee and S. Noguchi. Calcium dyshomeostasis in tubular aggregate myopathy. *International Journal of Molecular Sciences*, 17(11), NOV 2016. ISSN 1422-0067. doi: {10.3390/ijms17111952}.

[73] A. Mallick and A. R. Bodenham. MDMA induced hyperthermia: a survivor with an initial body temperature of 42.9 degrees C. *Journal of Accident & Emergency Medicine*, 14(5):336–338, SEP 1997. ISSN 1351-0622.

[74] C. K. Dao, S. M. Nowinski, and E. M. Mills. The heat is on: Molecular mechanisms of drug-induced hyperthermia. *Temperature*, 1(3):183–191, 2014. doi: 10.4161/23328940.2014.985953. URL `https://doi.org/10.4161/23328940.2014.985953`.

[75] E. M. Mills, M. L. Banks, J. E. Sprague, and T. Finkel. Pharmacology: Uncoupling the agony from ecstasy - A mitochondrial protein may mediate a dangerous side-effect of some recreational drugs. *Nature*, 426(6965):403–404, NOV 27 2003. ISSN 0028-0836. doi: {10.1038/426403a}.

[76] O. M. Kelly, Y. M. McNamara, L. H. Manzke, M. J. Meegan, and R. K. Porter. The preservation of in vivo phosphorylated and activated uncoupling protein 3 (UCP3) in isolated skeletal muscle mitochondria following administration of 3,4-methylenedioxymethamphetamine (MDMA aka ecstasy) to rats/mice. *Mitochondrion*, 12(1, SI):110–119, JAN 2012. ISSN 1567-7249. doi: {10.1016/j.mito.2011.03.011}.

[77] M. J. Bamshad, S. B. Ng, A. W. Bigham, H. K. Tabor, M. J. Emond, D. A. Nickerson, and J. Shendure. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, 12(11):745–755, NOV 2011. ISSN 1471-0056. doi: {10.1038/nrg3031}.

[78] C. Willyard. Expanded human gene tally reignites debate. *Nature*, 558(7710):354–355, JUN 21 2018. ISSN 0028-0836. doi: {10.1038/d41586-018-05462-w}.

[79] G. Bernardi. The genomic code: a pervasive encoding/molding of chromatin structures and a solution of the "non-coding DNA" mystery. *Bioessays*, 41 (12), DEC 2019. ISSN 0265-9247. doi: {10.1002/bies.201900106}.

[80] A. M. Meynert, M. Ansari, D. R. FitzPatrick, and M. S. Taylor. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics*, 15, JUL 19 2014. ISSN 1471-2105. doi: {10.1186/1471-2105-15-247}.

[81] A. Belkadi, A. Bolze, Y. Itan, A Cobat, Q. B. Vincent, A. Antipenko, L. Shang, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National*

Academy of Sciences of the United States of America, 112(17):5473–5478, APR 28 2015. ISSN 0027-8424. doi: {10.1073/pnas.1418631112}.

[82] J. Meienberg, R. Bruggmann, K. Oexle, and G. Matyas. Clinical sequencing: is WGS the better WES? *Human Genetics*, 135(3):359–362, MAR 2016. ISSN 0340-6717. doi: {10.1007/s00439-015-1631-9}.

[83] Y. Barbitoff, D. Polev, I. Shcherbakova, A. Kiselev, A. Glotov, E. Serebryakova, A. Kostareva, et al. Systematic dissection of biases in whole exome and whole genome strategies for human genome resequencing reveals major determinants of human coding sequence coverage. *European Journal of Human Genetics*, 27(1):509, JUL 2019. ISSN 1018-4813. 51st Conference of the European-Society-of-Human-Genetics (ESHG) in conjunction with the European Meeting on Psychosocial Aspects of Genetics (EMPAG), Milan, ITALY, JUN 16-19, 2018.

[84] C. F. Wright, D. R. FitzPatrick, and H. V. Firth. Paediatric genomics: diagnosing rare disease in children. *Nature Reviews Genetics*, 19(5):253–268, MAY 2018. ISSN 1471-0056. doi: {10.1038/nrg.2017.116}.

[85] V. Heinrich, J. Stange, T. Dickhaus, P. Imkeller, U. Krueger, S. Bauer, S. Mundlos, et al. The allele distribution in next-generation sequencing data sets is accurately described as the result of a stochastic branching process. *Nucleic Acids Research*, 40(6):2426–2431, MAR 2012. ISSN 0305-1048. doi: {10.1093/nar/gkr1073}.

[86] M. J. Clark, R. Chen, H. Y. K. Lam, K. J. Karczewski, R. Chen, G. Euskirchen, A. J. Butte, et al. Exome sequencing reveals novel rare variants in the ryanodine receptor and calcium channel genes in malignant hyperthermia familiesPerformance comparison of exome DNA sequencing technologies. *Nature Biotechnology*, 29(10):908–U206, OCT 2011. ISSN 1087-0156. doi: {10.1038/nbt.1975}.

[87] D. S. Marchuk, K. Crooks, N. Strande, K. Kaiser-Rogers, L. Milko, A. Brandt, A. Arreola, et al. Increasing the diagnostic yield of exome sequencing by copy number variant analysis. *PLOS One*, 13(12), DEC 17 2018. ISSN 1932-6203. doi: {10.1371/journal.pone.0209185}.

[88] R. Pfundt, M. del Rosario, L. E. L. M. Vissers, M. P. Kwint, I. M. Janssen, N. de Leeuw, H. G. Yntema, et al. Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. *Genetics in Medicine*, 19(6):667–675, JUN 2017. ISSN 1098-3600. doi: {10.1038/gim.2016.163}.

[89] A. Warr, C. Robert, D. Hume, A. Archibald, N. Deeb, and M. Watson. Exome sequencing: current and future perspectives. *G3- Genes Genomes Genetics*, 5(8):1543–1550, AUG 1 2015. ISSN 2160-1836. doi: {10.1534/g3.115.018564}.

[90] C. Sekhar R. Chilamakuri, S. Lorenz, M. Madoui, D. Vodak, J. Sun, E. Hovig, and others. Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*, 15, JUN 9 2014. ISSN 1471-2164. doi: {10.1186/1471-2164-15-449}.

[91] D. Sims, I. Sudbery, N. E. Ilott, A. Heger, and C. P. Ponting. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2):121–132, FEB 2014. ISSN 1471-0056. doi: {10.1038/nrg3642}.

[92] H. Li and R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760, JUL 15 2009. ISSN 1367-4803. doi: {10.1093/bioinformatics/btp324}.

[93] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–U54, APR 2012. ISSN 1548-7091. doi: {10.1038/NMETH.1923}.

[94] J. Chen, X. Li, H. Zhong, Y. Meng, and H. Du. Systematic comparison of germline variant calling pipelines cross multiple next-generation sequencers. *Scientific Reports*, 9, JUN 27 2019. ISSN 2045-2322. doi: {10.1038/s41598-019-45835-3}.

[95] S. Kim, K. Scheffler, A. L. Halpern, M. A. Bekritsky, E. Noh, M. Kallberg, X. Chen, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods*, 15(8):591+, AUG 2018. ISSN 1548-7091. doi: {10.1038/s41592-018-0051-x}.

[96] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, AUG 15 2009. ISSN 1367-4803. doi: {10.1093/bioinformatics/btp352}.

[97] Erik Garrison and Gabor Marth. Haplotype-based variant detection from short-read sequencing, 2012.

[98] J. O'Rawe, T. Jiang, G. Sun, Y. Wu, W. Wang, J. Hu, P. Bodily, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine*, 5, MAR 27 2013. ISSN 1756-994X. doi: {10.1186/gm432}.

[99] S. Hwang, E. Kim, I. Lee, and E. M. Marcotte. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific Reports*, 5, DEC 7 2015. ISSN 2045-2322. doi: {10.1038/srep17875}.

[100] M. Kumaran, U. Subramanian, and B. Devarajan. Performance assessment of variant calling pipelines using human whole exome sequencing and simulated data. *BMC Bioinformatics*, 20, JUN 17 2019. ISSN 1471-2105. doi: {10.1186/s12859-019-2928-9}.

[101] F. Cunningham, P. Achuthan, W. Akanni, J. Allen, M. R. Amode, I. M. Armean, R. Bennett, et al. Ensembl 2019. *Nucleic Acids Research*, 47(D1): D745–D751, JAN 8 2019. ISSN 0305-1048. doi: {10.1093/nar/gky1113}.

[102] N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. H. R. McVeigh, B. Rajput, B. Robbertse, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1):D733–D745, JAN 4 2016. ISSN 0305-1048. doi: {10.1093/nar/gkv1189}.

[103] D. J. McCarthy, P. Humburg, A. Kanapin, M. A. Rivas, K. Gaulton, J. Cazier, P. Donnelly, et al. Choice of transcripts and software has a large effect on variant annotation. *Genome Medicine*, 6, MAR 31 2014. ISSN 1756-994X. doi: {10.1186/gm543}.

[104] K. Wang, M. Li, and H. Hakonarson. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids*

Research, 38(16):e164–e164, 07 2010. ISSN 0305-1048. doi: 10.1093/nar/gkq603. URL `https://doi.org/10.1093/nar/gkq603`.

[105] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, et al. The Ensembl variant effect predictor. *Genome Biology*, 17, JUN 6 2016. ISSN 1474-760X. doi: {10.1186/s13059-016-0974-4}.

[106] P. Cingolani, A. Platts, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2):80–92, 2012.

[107] L. Clarke, S. Fairley, X. Zheng-Bradley, I. Streeter, E. Perry, E. Lowy, A. Tasse, et al. The international genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Research*, 45(D1):D854–D859, JAN 4 2017. ISSN 0305-1048. doi: {10.1093/nar/gkw829}.

[108] K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 2019. doi: 10.1101/531210. URL `https://www.biorxiv.org/content/early/2019/08/13/531210`.

[109] M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616):285+, AUG 18 2016. ISSN 0028-0836. doi: {10.1038/nature19057}.

[110] D. M. Altshuler, R. M. Durbin, G. R. Abecasis, D. R. Bentley, A. Chakravarti, A. G. Clark, P. Donnelly, et al. A global reference for human genetic variation. *Nature*, 526(7571):68+, OCT 1 2015. ISSN 0028-0836. doi: {10.1038/nature15393}.

[111] J. D. Wall, E. W. Stawiski, A. Ratan, H. L. Kim, C. Kim, R. Gupta, K. Suryamohan, et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*, 576(7785):106+, DEC 5 2019. ISSN 0028-0836. doi: {10.1038/s41586-019-1793-z}.

[112] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, and M. Kircher. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1):D886–D894, JAN 8 2019. ISSN 0305-1048. doi: {10.1093/nar/gky1016}.

[113] N. M. Ioannidis, J. H. Rothstein, V. Pejaver, S. Middha, S. K. McDonnell, S. Baheti, A. Musolf, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *American Journal of Human Genetics*, 99(4):877–885, OCT 6 2016. ISSN 0002-9297. doi: {10.1016/j.ajhg.2016.08.016}.

[114] A. H. Schiemann and K. M. Stowell. Comparison of pathogenicity prediction tools on missense variants in RYR1 and CACNA1S associated with malignant hyperthermia. *British Journal of Anaesthesia*, 117(1):124–128, JUL 2016. ISSN 0007-0912. doi: {10.1093/bja/aew065}.

[115] J. Li, T. Zhao, Y. Zhang, K. Zhang, L. Shi, Y. Chen, X. Wang, et al. Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Research*, 46(15):7793–7804, SEP 6 2018. ISSN 0305-1048. doi: {10.1093/nar/gky678}.

[116] S. Carbon, E. Douglass, N. Dunn, B. Good, N. L. Harris, S. E. Lewis, C. J. Mungall, et al. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1):D330–D338, JAN 8 2019. ISSN 0305-1048. doi: {10.1093/nar/gky1055}.

[117] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, et al. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, MAY 2000. ISSN 1061-4036. doi: {10.1038/75556}.

[118] J. S. Amberger, C. A. Bocchini, F. Schiettecatte, A. F. Scott, and A. Hamosh. OMIM.org: Online Mendelian Inheritance in Man (OMIM (R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Research*, 43(D1):D789–D798, JAN 28 2015. ISSN 0305-1048. doi: {10.1093/nar/gku1205}.

[119] S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin. dbSNP: the NCBI database of genetic variation. *Nucleic*

Acids Research, 29(1):308–311, JAN 1 2001. ISSN 0305-1048. doi: {10.1093/nar/29.1.308}.

[120] J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, et al. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6):580–585, JUN 2013. ISSN 1061-4036. doi: {10.1038/ng.2653}.

[121] M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, et al. Tissue-based map of the human proteome. *Science*, 347(6220), 2015. ISSN 0036-8075. doi: 10.1126/science.1260419. URL `https://science.sciencemag.org/content/347/6220/1260419`.

[122] A. N. Haddock, S. A. Labuzan, A. E. Haynes, C. S. Hayes, K. M. Kakareka, and D. S. Waddell. Dual-specificity phosphatase 4 is upregulated during skeletal muscle atrophy and modulates extracellular signal-regulated kinase activity. *American Journal of Physiology- Cell Physiology*, 316(4):C567–C581, APR 2019. ISSN 0363-6143. doi: {10.1152/ajpcell.00234.2018}.

[123] J. Desvignes, M. Bartoli, V. Delague, M. Krahn, M. Miltgen, C. Broud, and D. Salgado. VarAFT: a variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Research*, 46(W1):W545–W553, 05 2018. ISSN 0305-1048. doi: 10.1093/nar/gky471. URL `https://doi.org/10.1093/nar/gky471`.

[124] U. Paila, B. A. Chapman, R. Kirchner, and A. R. Quinlan. GEMINI: integrative exploration of genetic variation and genome annotations. *PLOS Computational Biology*, 9(7), JUL 2013. doi: {10.1371/journal.pcbi.1003153}.

[125] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1):D457–D462, JAN 4 2016. ISSN 0305-1048. doi: {10.1093/nar/gkv1070}.

[126] E. Talevich, A. H. Shain, T Botton, and B. C. Bastian. Cnvkit: Genome-wide copy number detection and visualization from targeted dna sequencing. *PLOS Computational Biology*, 12(4):1–18, 04 2016. doi: 10.1371/

journal.pcbi.1004873. URL https://doi.org/10.1371/journal.pcbi.1004873.

[127] Y. Jiang, R. Wang, E. Urrutia, I. N. Anastopoulos, K. L. Nathanson, and N. R. Zhang. CODEX2: full-spectrum copy number variation detection by high-throughput DNA sequencing. *Genome Biology*, 19, NOV 26 2018. ISSN 1474-760X. doi: {10.1186/s13059-018-1578-y}.

[128] S. Mangul, T. Mosqueiro, R. J. Abdill, D. Duong, K. Mitchell, V. Sarwal, B. Hill, et al. Challenges and recommendations to improve the installability and archival stability of omics computational tools. *PLOS Biology*, 17(6), JUN 2019. ISSN 1544-9173. doi: {10.1371/journal.pbio.3000333}.

[129] S. Nambot, J. Thevenon, P. Kuentz, Y. Duffourd, E. Tisserant, A. Bruel, A. Mosca-Boidron, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genetics in Medicine*, 20(6):645–654, JUN 2018. ISSN 1098-3600. doi: {10.1038/gim.2017.162}.

[130] R. L. Brown, A. N. Pollock, K. G. Couchman, M. Hodges, D. O. Hutchinson, R. Waaka, P. Lynch, et al. A novel ryanodine receptor mutation and genotype-phenotype correlation in a large malignant hyperthermia New Zealand Maori pedigree. *Human Molecular Genetics*, 9(10):1515–1524, JUN 12 2000. ISSN 0964-6906. doi: {10.1093/hmg/9.10.1515}.

[131] S. Richards, N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5):405–424, MAY 2015. ISSN 1098-3600. doi: {10.1038/gim.2015.30}.

[132] P. Kaliman and E. Llagostera. Myotonic dystrophy protein kinase (DMPK) and its role in the pathogenesis of myotonic dystrophy 1. *Cellular Signalling*, 20(11):1935–1941, NOV 2008. ISSN 0898-6568. doi: {10.1016/j.cellsig.2008.05.005}.

[133] T. Kimura, M. Nakamori, J. D. Lueck, P. Pouliquin, F. Aoike, H. Fujimura,

R. T. Dirksen, et al. Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase in myotonic dystrophy type 1. *Human Molecular Genetics*, 14(15):2189–2200, AUG 1 2005. ISSN 0964-6906. doi: {10.1093/hmg/ddi223}.

[134] Picard toolkit. `http://broadinstitute.github.io/picard/`, 2019.

[135] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303, SEP 2010. ISSN 1088-9051. doi: {10.1101/gr.107524.110}.

[136] H. Wickham. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

[137] B. S. Pedersen and A. R. Quinlan. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, 34(5):867–868, 10 2017. ISSN 1367-4803. doi: 10.1093/bioinformatics/btx699. URL `https://doi.org/10.1093/bioinformatics/btx699`.

[138] D. H. Phanstiel, A. P. Boyle, C. L. Araya, and M. P. Snyder. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics*, 30(19):2808–2810, 06 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu379. URL `https://doi.org/10.1093/bioinformatics/btu379`.

[139] I. Milne, G. Stephen, M. Bayer, P. J. A. Cock, L. Pritchard, L. Cardle, P. D. Shaw, et al. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, 14(2, SI):193–202, MAR 2013. ISSN 1467-5463. doi: {10.1093/bib/bbs012}.

[140] E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Cech, J. Chilton, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1):W537–W544, JUL 2 2018. ISSN 0305-1048. doi: {10.1093/nar/gky379}.

[141] A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, MAR 15 2010. ISSN 1367-4803. doi: {10.1093/bioinformatics/btq033}.

[142] G. Bonne, F. Rivier, and D. Hamroun. The 2018 version of the gene table of monogenic neuromuscular disorders (nuclear genome). *Neuromuscular Disorders*, 27(12):1152–1183, DEC 2017. ISSN 0960-8966. doi: {10.1016/j.nmd.2017.10.005}.

[143] F. Helmbacher. Tissue-specific activities of the fat1 cadherin cooperate to control neuromuscular morphogenesis. *PLOS Biology*, 16(5):1–54, 05 2018. doi: 10.1371/journal.pbio.2004734. URL `https://doi.org/10.1371/journal.pbio.2004734`.

[144] N. Caruso, B. Herberth, M. Bartoli, F. Puppo, J. Dumonceaux, A. Zimmermann, S. Denadai, et al. Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy. *PLOS Genetics*, 9(6), JUN 2013. ISSN 1553-7404. doi: {10.1371/journal.pgen.1003550}.

[145] V. Martinez-Redondo, A. T. Pettersson, and J. L. Ruas. The hitchhiker's guide to PGC-1 alpha isoform structure and biological functions. *Diabetologia*, 58(9):1969–1977, SEP 2015. ISSN 0012-186X. doi: {10.1007/s00125-015-3671-z}.

[146] A. Tournadre, V. Lenief, and P. Miossec. Expression of Toll-like receptor 3 and Toll-like receptor 7 in muscle is characteristic of inflammatory myopathy and is differentially regulated by Th1 and Th17 cytokines. *Arthritis and Rheumatism*, 62(7):2144–2151, JUL 2010. ISSN 0004-3591. doi: {10.1002/art.27465}.

[147] B. Schreiner, J. Voss, J. R. Wischhusen, Y. Dombrowski, A. Steinle, H. Lochmuller, M. Dalakas, et al. Expression of toll-like receptors by human muscle cells in vitro and in vivo: TLR3 is highly expressed in inflammatory and HIV myopathies, mediates IL-8 release, and up-regulation of NKG2D-ligands. *Faseb Journal*, 19(13):118+, NOV 2005. ISSN 0892-6638. doi: {10.1096/fj.05-4342fje}.

[148] M. C. Towler and G. D. Hardie. AMP-activated protein kinase in metabolic control and insulin signaling. *Circulation Research*, 100(3):328–341, FEB 16 2007. ISSN 0009-7330. doi: {10.1161/01.RES.0000256090.42690.05}.

[149] H. Koepsell. Organic cation transporters in health and disease. *Pharmacological Reviews*, 72(1):253–319, JAN 2020. ISSN 0031-6997. doi: {10.1124/pr.118.015578}.

[150] A. L. Anding, C. Wang, T. Chang, D. A. Sliter, C. M Powers, K. Hofmann, R. J. Youle, and E. H. Baehrecke. Vps13d encodes a ubiquitin-binding protein that is required for the regulation of mitochondrial size and clearance. *Current Biology*, 28(2):287 – 295.e6, 2018. ISSN 0960-9822. doi: https://doi.org/10.1016/j.cub.2017.11.064. URL `http://www.sciencedirect.com/science/article/pii/S0960982217315890`.

[151] A. M. DeSimone, A. Pakula, A. Lek, and Emerson Jr. C. P. *Facioscapulohumeral muscular dystrophy*, pages 1229–1279. American Cancer Society, 2017. ISBN 9780470650714. doi: 10.1002/cphy.c160039. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/cphy.c160039`.

[152] R. Tawil and S. M. Van der Maarel. Facioscapulohumeral muscular dystrophy. *Muscle & Nerve*, 34(1):1–15, JUL 2006. ISSN 0148-639X. doi: {10.1002/mus.20522}.

[153] X. Wu, R. Kekuda, W. Huang, Y. J. Fei, F. H. Leibach, J. W. Chen, S. J. Conway, et al. Identity of the organic cation transporter OCT3 as the extraneuronal monoamine transporter (uptake(2)) and evidence for the expression of the transporter in the brain. *Journal Of Biological Chemistry*, 273(49):32776–32786, DEC 4 1998. ISSN 0021-9258. doi: {10.1074/jbc.273.49.32776}.

[154] C. Dobson-Stone, A. Velayos-Baeza, A. Jansen, F. Andermann, F. Dubeau, F. Robert, A. Summers, et al. Identification of a VPS13A founder mutation in French Canadian families with chorea-acanthocytosis. *Neurogenetics*, 6(3):151–158, SEP 2005. ISSN 1364-6745. doi: {10.1007/s10048-005-0220-9}.

[155] J. D. Lin, P. Puigserver, J. Donovan, P. Tarr, and B. M. Spiegelman. Peroxisome proliferator-activated receptor gamma coactivator 1 beta (PGC-1

beta), a novel PGC-1-related transcription coactivator associated with host cell factor. *Journal Of Biological Chemistry*, 277(3):1645–1648, JAN 18 2002. ISSN 0021-9258. doi: {10.1074/jbc.C100631200}.

[156] T. Kawai and S. Akira. TLR signaling. *Cell Death and Differentiation*, 13 (5):816–825, MAY 2006. ISSN 1350-9047. doi: {10.1038/sj.cdd.4401850}.

[157] C. P. Hodgkinson, R. E. Pratt, I. Kirste, S. Dal-Pra, J. P. Cooke, and V. J. Dzau. Cardiomyocyte maturation requires tlr3 activated nuclear factor Kappa b. *Stem Cells*, 36(8):1198–1209, AUG 2018. ISSN 1066-5099. doi: {10.1002/stem.2833}.

[158] X. Gao, S. Gao, Y. Guan, L. Huang, J. Huang, L. Lin, Y. Liu, et al. Toll-like receptor 3 controls QT interval on the electrocardiogram by targeting the degradation of Kv4.2/4.3 channels in the endoplasmic reticulum. *Faseb Journal*, 33(5):6197–6208, MAY 2019. ISSN 0892-6638. doi: {10.1096/fj.201801464R}.

[159] D. Garcia and R. J. Shaw. Ampk: mechanisms of cellular energy sensing and restoration of metabolic balance. *Molecular Cell*, 66(6):789–800, JUN 15 2017. ISSN 1097-2765. doi: {10.1016/j.molcel.2017.05.032}.

[160] S. L. McGee, K. F. Howlett, R. L. Starkie, D. Cameron-Smith, B. E. Kemp, and M. Hargreaves. Exercise increases nuclear AMPK alpha(2) in human skeletal muscle. *Diabetes*, 52(4):926–928, APR 2003. ISSN 0012-1797. doi: {10.2337/diabetes.52.4.926}.

[161] S. Maerki, M. H. Olma, T. Staubli, P. Steigemann, D. W. Gerlich, M. Quadroni, I. Sumara, et al. The Cul3-KLHL21 E3 ubiquitin ligase targets Aurora B to midzone microtubules in anaphase and is required for cytokinesis. *Journal of Cell Biology*, 187(6):791–800, DEC 14 2009. ISSN 0021-9525. doi: {10.1083/jcb.200906117}.

[162] J. M. Holaska, S. Rais-Bahrami, and K. L. Wilson. Lmo7 is an emerin-binding protein that regulates the transcription of emerin and many other muscle-relevant genes. *Human Molecular Genetics*, 15(23):3459–3472, DEC 2006. ISSN 0964-6906. doi: {10.1093/hmg/ddl423}.

[163] A. Marg, H. Haase, T. Neumann, M. Kouno, and I. Morano. AHNAK1 and AHNAK2 are costameric proteins: AHNAK1 affects transverse skeletal muscle fiber stiffness. *Biochemical and Biophysical Research Communications*, 401(1):143–148, OCT 8 2010. ISSN 0006-291X. doi: {10.1016/j.bbrc.2010.09.030}.

[164] M. S. C. Wilson, T. M. Livermore, and A. Saiardi. Inositol pyrophosphates: between signalling and metabolism. *Biochemical Journal*, 452(3):369–379, JUN 15 2013. ISSN 0264-6021. doi: {10.1042/BJ20130118}.

[165] A. Bateman, M. Martin, S. Orchard, M. Magrane, E. Alpi, B. Bely, M. Bingley, et al. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1):D506–D515, JAN 8 2019. ISSN 0305-1048. doi: {10.1093/nar/gky1049}.

[166] W. Cheng, F. Yang, C. L. Takanishi, and J. Zheng. Thermosensitive TRPV channel subunits coassemble into heteromeric channels with intermediate conductance and gating properties. *Journal of General Physiology*, 129(3): 191–207, MAR 2007. ISSN 0022-1295. doi: {10.1085/jgp.200709731}.

[167] V. A. Gupta and A. H. Beggs. Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skeletal Muscle*, 4, JUN 1 2014. ISSN 2044-5040. doi: {10.1186/2044-5040-4-11}.

[168] G. Bonne, M. R. Di Barletta, S. Varnous, H. M. Becane, E. H. Hammouda, L. Merlini, F. Muntoni, et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nature Genetics*, 21(3):285–288, MAR 1999. ISSN 1061-4036. doi: {10.1038/6799}.

[169] A. Mull, G. Kim, and J. M. Holaska. LMO7-null mice exhibit phenotypes consistent with Emery-Dreifuss muscular dystrophyY. *Muscle & Nerve*, 51 (2):222–228, FEB 2015. ISSN 0148-639X. doi: {10.1002/mus.24286}.

[170] D. H. Lao, M. C. Esparza, S. N. Bremner, I. Banerjee, J. Zhang, J. Veevers, W. H. Bradford, et al. Lmo7 is dispensable for skeletal muscle and cardiac function. *American Journal of Physiology- Cell Physiology*, 309(7):C470–C479, OCT 1 2015. ISSN 0363-6143. doi: {10.1152/ajpcell.00177.2015}.

[171] A. Komuro, Y. Masuda, K. Kobayashi, R. Babbitt, M. Gunel, R. A. Flavell, and V. T. Marchesi. The AHNAKs are a class of giant propeller-like proteins that associate with calcium channel proteins of cardiornyocytes and other cells. *Proceedings of the National Academy of Sciences of the United States of America*, 101(12):4053–4058, MAR 23 2004. ISSN 0027-8424. doi: {10.1073/pnas.0308619101}.

[172] I. Pankonien, J. L. Alvarez, A. Doller, C. Koehncke, D. Rotte, V. Regitz-Zagrosek, I. Morano, et al. Ahnak1 is a tuneable modulator of cardiac Ca(v)1.2 calcium channel activity. *Journal of Muscle Research and Cell Motility*, 32(4-5, SI):281–290, DEC 2011. ISSN 0142-4319. doi: {10.1007/s10974-011-9269-2}.

[173] E. L. Huttlin, R. J. Bruckner, J. A. Paulo, J. R. Cannon, L. Ting, K. Baltier, G. Colby, et al. Architecture of the human interactome defines protein communities and disease networks. *Nature*, 545(7655):505+, MAY 25 2017. ISSN 0028-0836. doi: {10.1038/nature22366}.

[174] D. Pareyson and C. Marchesi. Diagnosis, natural history, and management of Charcot-Marie-Tooth disease. *Lancet Neurology*, 8(7):654–667, JUL 2009. ISSN 1474-4422. doi: {10.1016/S1474-4422(09)70110-3}.

[175] J. W. Kim, G. Kim, T. W. Kim, W. Han, J. H. Maeng, C. Y. Jeong, J. H. Choi, et al. Anesthesia in a patient with Charcot-Marie-Tooth disease with pneumothorax: a case report. *Journal of International Medical Research*, 47(11):5896–5902, NOV 2019. ISSN 0300-0605. doi: {10.1177/0300060519881239}.

[176] M. L. Kennerson, N. T. Nassif, J. L. Dawkins, R. M. DeKroon, J. G. Yang, and G. A. Nicholson. The Charcot-Marie-Tooth binary repeat contains a gene transcribed from the opposite strand of a partially duplicated region of the COX10 gene. *Genomics*, 46(1):61–69, NOV 15 1997. ISSN 0888-7543. doi: {10.1006/geno.1997.5012}.

[177] A. Chakraborty, S. Kim, and S. H. Snyder. Inositol pyrophosphates as mammalian cell signals. *Science Signaling*, 4(188), AUG 30 2011. ISSN 1945-0877. doi: {10.1126/scisignal.2001958}.

[178] S. B. Shears. Intimate connections: Inositol pyrophosphates at the interface of metabolic regulation and cell signaling. *Journal of Cellular Physiology*, 233(3):1897–1912, MAR 2018. ISSN 0021-9541. doi: {10.1002/jcp.26017}.

[179] S. Hauke, A. K. Dutta, V. B. Eisenbeis, D. Bezold, T. Bittner, C. Wittwer, D. Thakor, et al. Photolysis of cell-permeant caged inositol pyrophosphates controls oscillations of cytosolic calcium in a beta-cell line. *Chemical Science*, 10(9):2687–2692, MAR 7 2019. ISSN 2041-6520. doi: {10.1039/c8sc03479f}.

[180] P. C. Fridy, J. C. Otto, D. E. Dollins, and J. D. York. Cloning and characterization of two human VIP1-like inositol hexakisphosphate and diphosphoinositol pentakisphosphate kinases. *Journal Of Biological Chemistry*, 282(42):30754–30762, OCT 19 2007. doi: {10.1074/jbc.M704656200}.

[181] F. Vanden Abeele, S. Lotteau, S. Ducreux, C. Dubois, N. Monnier, A. Hanna, D. Gkika, et al. TRPV1 variants impair intracellular Ca^{2+} signaling and may confer susceptibility to malignant hyperthermia. *Genetics in Medicine*, 21(2):441–450, FEB 2019. ISSN 1098-3600. doi: {10.1038/s41436-018-0066-9}.

[182] G. E. Torres, R. R. Gainetdinov, and M. G. Caron. Plasma membrane monoamine transporters: Structure, regulation and function. *Nature Reviews Neuroscience*, 4(1):13–25, JAN 2003. ISSN 1471-003X. doi: {10.1038/nrn1008}.

[183] J. E. Lawson, S. H. Park, A. R. Mattison, J. G. Yan, and L. J. Reed. Cloning, expression, and properties of the regulatory subunit of bovine pyruvate dehydrogenase phosphatase. *Journal Of Biological Chemistry*, 272 (50):31625–31629, DEC 12 1997. ISSN 0021-9258. doi: {10.1074/jbc.272.50.31625}.

[184] F. B. Berry, Y. Miura, K. Mihara, P. Kaspar, N. Sakata, T. Hashimoto-Tamaoki, and T. Tamaoki. Positive and negative regulation of myogenic differentiation of C2C12 cells by isoforms of the multiple homeodomain zinc finger transcription factor ATBF1. *Journal Of Biological Chemistry*, 276(27):25057–25065, JUL 6 2001. ISSN 0021-9258. doi: {10.1074/jbc.M010378200}.

[185] Y. Zhong, L. Zou, Z. Wang, Y. Pan, Z. Dai, X. Liu, L. Cui, et al. Lrrc75b is a novel negative regulator of C2C12 myogenic differentiation. *International Journal of Molecular Medicine*, 38(5):1411–1418, NOV 2016. ISSN 1107-3756. doi: {10.3892/ijmm.2016.2738}.

[186] K. Nonogaki. New insights into sympathetic regulation of glucose and fat metabolism. *Diabetologia*, 43(5):533–549, MAY 2000. ISSN 0012-186X. doi: {10.1007/s001250051341}.

[187] Dan Z., Xueying H., Lili H., Jianpeng W., Xi Z., Ju-Hong J., and Qiang Z. Transcription factor zfhx3 regulates calcium influx in mammary epithelial cells in part via the trpv6 calcium channel. *Biochemical and Biophysical Research Communications*, 519(2):366 – 371, 2019. ISSN 0006-291X. doi: https://doi.org/10.1016/j.bbrc.2019.08.148. URL `http://www.sciencedirect.com/science/article/pii/S0006291X19316778`.

[188] Z. Hu and D. Zou. Genotype-phenotype associations in atrial fibrillation: meta-analysis. *Journal of Interventional Cardiac Electrophysiology*, 54(3): 283–288, APR 2019. ISSN 1383-875X. doi: {10.1007/s10840-018-0484-2}.

[189] Y. Kao, J. Hsu, Y. Chen, Y. Lin, B. Lkhagva, S. Chen, and Y. Chen. ZFHX3 knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes. *International Journal of Cardiology*, 210:85–92, MAY 1 2016. ISSN 0167-5273. doi: {10.1016/j.ijcard.2016.02.091}.

[190] A. S. McElhinny, C. Schwach, M. Valichnac, S. Mount-Patrick, and C. C. Gregorio. Nebulin regulates the assembly and lengths of the thin filaments in striated muscle. *Journal of Cell Biology*, 170(6):947–957, SEP 12 2005. ISSN 0021-9525. doi: {10.1083/jcb.200502158}.

[191] S. Rousset, M. C. Alves-Guerra, J. Mozo, B. Miroux, A. M. Cassard-Doulcier, F. Bouillaud, and D. Ricquier. The biology of mitochondrial uncoupling proteins. *Diabetes*, 53(1):S130–S135, FEB 2004. ISSN 0012-1797. doi: {10.2337/diabetes.53.2007.S130}. 4th Servier-IGIS Symposium, ST JEAN CAP FERRAT, FRANCE, MAR 20-23, 2003.

[192] N. Sambuughin, W. Swietnicki, S. Techtmann, V. Matrosova, T. Wallace, L. Goldfarb, and E. Maynard. KBTBD13 interacts with Cullin 3 to

form a functional ubiquitin ligase. *Biochemical and Biophysical Research Communications*, 421(4):743–749, MAY 18 2012. ISSN 0006-291X. doi: {10.1016/j.bbrc.2012.04.074}.

[193] C. A. Sewry, J. M. Laitila, and C. Wallgren-Pettersson. Nemaline myopathies: a current view. *Journal of Muscle Research and Cell Motility*, 40(2, SI):111–126, JUN 2019. ISSN 0142-4319. doi: {10.1007/s10974-019-09519-9}. European Muscle Conference (EMC) / 48th Annual Meeting of the European-Society-for-Muscle-Research (ESMR), Univ Kent, Canterbury, ENGLAND, SEP 07-11, 2019.

[194] N. H. Tran and D. Smith. Anesthetic consideration for patients with nemaline rod myopathy: a literature review. *Pediatric Anesthesia and Critical Care Journal*, 5(1):31–39, 2017. ISSN 2281-8421. doi: {10.14587/paccj.2017.5}.

[195] M. Garibaldi, F. Fattori, C. A. Bortolotti, G. Brochier, C. Labasse, M. Verardo, E. Servian-Morilla, et al. Core-rod myopathy due to a novel mutation in BTB/POZ domain of KBTBD13 manifesting as late onset LGMD. *Acta Neuropathologica Communications*, 6, SEP 13 2018. ISSN 2051-5960. doi: {10.1186/s40478-018-0595-0}.

[196] N. Sambuughin, K. S. Yau, M. Olive, R. M. Duff, M. Bayarsaikhan, S. Lu, L. Gonzalez-Mera, et al. Dominant mutations in KBTBD13, a member of the BTB/Kelch family, cause nemaline myopathy with cores. *American Journal of Human Genetics*, 87(6):842–847, DEC 10 2010. ISSN 0002-9297. doi: {10.1016/j.ajhg.2010.10.020}.

[197] A. VidalPuig, G. Solanes, D. Grujic, J. S. Flier, and B. B. Lowell. UCP3: An uncoupling protein homologue expressed preferentially and abundantly in skeletal muscle and brown adipose tissue. *Biochemical and Biophysical Research Communications*, 235(1):79–82, JUN 9 1997. ISSN 0006-291X. doi: {10.1006/bbrc.1997.6740}.

[198] P. R. Cox and H. Y. Zoghbi. Sequencing, expression analysis, and mapping of three unique human tropomodulin genes and their mouse orthologs. *Genomics*, 63(1):97–107, JAN 1 2000. ISSN 0888-7543. doi: {10.1006/geno.1999.6061}.

[199] T. Ramadan, S. M. R. Camargo, V. Summa, P. Hunziker, S. Chesnov, K. M. Pos, and F. Verrey. Basolateral aromatic amino acid transporter TAT1 (Slc16A10) functions as an efflux pathway. *Journal of Cellular Physiology*, 206(3):771–779, MAR 2006. ISSN 0021-9541. doi: {10.1002/jcp.20531}.

[200] Q. Zhang, C. Bethmann, N. F. Worth, J. D. Davies, C. Wasner, A. Feuer, C. D. Ragnauth, et al. Nesprin-1 and -2 are involved in the pathogenesis of Emery-Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Human Molecular Genetics*, 16(23):2816–2833, DEC 1 2007. ISSN 0964-6906. doi: {10.1093/hmg/ddm238}.

[201] M. Baumann, E. Steichen-Gersdorf, B. Krabichler, B. Petersen, U. Weber, W. M. Schmidt, J. Zschocke, et al. Homozygous SYNE1 mutation causes congenital onset of muscular weakness with distal arthrogryposis: a genotype-phenotype correlation. *European Journal of Human Genetics*, 25 (2):262–266, FEB 2017. ISSN 1018-4813. doi: {10.1038/ejhg.2016.144}.

[202] K. Kunzelmann, I. Cabrita, P. Wanitchakool, J. Ousingsawat, L. Sirianant, R. Benedetto, and R. Schreiber. Modulating Ca^{2+} signals: a common theme for TMEM16, Ist2, and TMC. *European Journal of Physiology*, 468(3):475–490, MAR 2016. ISSN 0031-6768. doi: {10.1007/s00424-015-1767-4}.

[203] M. Almannai, A. Alasmari, A. Alqasmi, E. Faqeih, F. Al Mutairi, M. Alotaibi, M. M. Samman, et al. Expanding the phenotype of SLC25A42-associated mitochondrial encephalomyopathy. *Clinical Genetics*, 93(5): 1097–1102, MAY 2018. ISSN 0009-9163. doi: {10.1111/cge.13210}.

[204] S. Tsutsumi, N. Kamata, T. J. Vokes, Y. Maruoka, K. Nakakuki, S. Enomoto, K. Omura, et al. The novel gene encoding a putative transmembrane protein is mutated in gnathodiaphyseal dysplasia (GDD). *American Journal of Human Genetics*, 74(6):1255–1261, JUN 2004. ISSN 0002-9297. doi: {10.1086/421527}.

[205] V. Bolduc, G. Marlow, K. M. Boycott, K. Saleki, H. Inoue, J. Kroon, M. Itakura, et al. Recessive mutations in the putative calcium-activated chloride channel anoctamin 5 cause proximal LGMD2L and distal MMD3 muscular dystrophies. *American Journal of Human Genetics*, 86(2):213–221, FEB 12 2010. ISSN 0002-9297. doi: {10.1016/j.ajhg.2009.12.013}.

[206] T. T. T. Phuong, J. An, S. H. Park, A. Kim, H. B. Choi, and T. M. Kang. Deficiency of anoctamin 5/TMEM16E causes nuclear positioning defect and impairs Ca^{2+} signaling of differentiated C2C12 myotubes. *Korean Journal of Physiology & Pharmacology*, 23(6):539–547, NOV 2019. ISSN 1226-4512. doi: {10.4196/kjpp.2019.23.6.539}.

[207] A. Dayal, S. F. J. Ng, and M. Grabner. Ca^{2+}-activated Cl- channel TMEM16A/ANO1 identified in zebrafish skeletal muscle is crucial for action potential acceleration. *Nature Communications*, 10, JAN 10 2019. ISSN 2041-1723. doi: {10.1038/s41467-018-07918-z}.

[208] O. Alekhina, E. Burstein, and D. D. Billadeau. Cellular functions of WASP family proteins at a glance. *Journal of Cell Science*, 130(14):2235–2241, JUL 15 2017. ISSN 0021-9533. doi: {10.1242/jcs.199570}.

[209] S. T. de Bot, S. Vermeer, W. Buijsman, A. Heister, M. Voorendt, A. Verrips, H. Scheffer, et al. Pure adult-onset Spastic Paraplegia caused by a novel mutation in the KIAA0196 (SPG8) gene. *Journal of Neurology*, 260(7): 1765–1769, JUL 2013. ISSN 0340-5354. doi: {10.1007/s00415-013-6870-x}.

[210] A. M. Elliott, L. R. Simard, G. Coghlan, A. E. Chudley, B. N. Chodirker, C. R. Greenberg, T. Burch, et al. A novel mutation in KIAA0196: identification of a gene involved in Ritscher-Schinzel/3C syndrome in a First Nations cohort. *Journal of Medical Genetics*, 50(12):819–822, DEC 2013. ISSN 0022-2593. doi: {10.1136/jmedgenet-2013-101715}.

[211] G. Singhania and A. Koratala. Hypokalemic periodic paralysis with mutations in WASHC5 and ANKH genes. *American Journal of Kidney Diseases*, 71(4):584, APR 2018. ISSN 0272-6386. Spring Clinical Meeting of the National-Kidney Foundation (NKD), Austin, TX, APR 10-14, 2018.

[212] F. Ropers, E. Derivery, H. Hu, M. Garshasbi, M. Karbasiyan, M. Herold, G. Nuernberg, et al. Identification of a novel candidate gene for non-syndromic autosomal recessive intellectual disability: the WASH complex member SWIP. *Human Molecular Genetics*, 20(13):2585–2590, JUL 1 2011. ISSN 0964-6906. doi: {10.1093/hmg/ddr158}.

[213] H. Zhang, A. Qiao, D. Yang, L. Yang, A. Dai, C. De Graaf, S. Reedtz-Runge, et al. Structure of the full-length glucagon class B G-protein-coupled

receptor. *Nature*, 546(7657):259+, JUN 8 2017. ISSN 0028-0836. doi: {10.1038/nature22363}.

[214] R. Artuso, F. T. Papa, E. Grillo, M. Mucciolo, D. H. Yasui, K. W. Dunaway, V. Disciglio, et al. Investigation of modifier genes within copy number variations in Rett syndrome. *Journal of Human Genetics*, 56(7):508–515, JUL 2011. ISSN 1434-5161. doi: {10.1038/jhg.2011.50}.

[215] R. Chettier, K. Ward, and H. M. Albertsen. Endometriosis is associated with rare copy number variants. *PLOS One*, 9(8), AUG 1 2014. ISSN 1932-6203. doi: {10.1371/journal.pone.0103968}.

[216] S. Ran, Y. Liu, L. Zhang, Y. Pei, T. Yang, R. Hai, Y. Han, et al. GGenome-wide association study identified copy number variants important for appendicular lean mass. *PLOS One*, 9(3), MAR 13 2014. ISSN 1932-6203. doi: {10.1371/journal.pone.0089776}.

[217] H. Hibino, A. Inanobe, K. Furutani, S. Murakami, I. Findlay, and Y. Kurachi. Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiological Reviews*, 90(1):291–366, JAN 2010. ISSN 0031-9333. doi: {10.1152/physrev.00021.2009}.

[218] R. M. Paninka, D. R. Mazzotti, M. M. L. Kizys, A. C. Vidi, H. Rodrigues, S. P. Silva, I. S. Kunii, et al. Whole genome and exome sequencing realignment supports the assignment of KCNJ12, KCNJ17, and KCNJ18 paralogous genes in thyrotoxic periodic paralysis locus: functional characterization of two polymorphic Kir2.6 isoforms. *Molecular Genetics and Genomics*, 291(4):1535–1544, AUG 2016. ISSN 1617-4615. doi: {10.1007/s00438-016-1185-0}.

[219] A. A. Anderson, S. Treves, D. Biral, R. Betto, D. Sandona, M. Ronjat, and F. Zorzato. The novel skeletal muscle sarcoplasmic reticulum JP-45 protein - Molecular cloning, tissue distribution, developmental expression, and interaction with alpha 1.1 subunit of the voltage-gated calcium channel. *Journal Of Biological Chemistry*, 278(41):39987–39992, OCT 10 2003. ISSN 0021-9258. doi: {10.1074/jbc.M305016200}.

[220] O. Delbono, J. Xia, S. Treves, Z. Wang, R. Jimenez-Moreno, A. M. Payne, M. L. Messi, et al. Loss of skeletal muscle strength by ablation of the

sarcoplasmic reticulum protein JP45. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):20108–20113, DEC 11 2007. ISSN 0027-8424. doi: {10.1073/pnas.0707389104}.

[221] T. Yasuda, O. Delbono, Z. Wang, M. L. Messi, T. Girard, A. Urwyler, S. Treves, et al. JP-45/JSRP1 variants affect skeletal muscle excitation-contraction coupling by decreasing the sensitivity of the dihydropyridine receptor. *Human Mutation*, 34(1):184–190, JAN 2013. ISSN 1059-7794. doi: {10.1002/humu.22209}.

[222] R. P. Haugland, J. Yguerabide, and L. Stryer. Dependence of the kinetics of singlet-singlet energy transfer on spectral overlap. *Proceedings of the National Academy of Sciences of the United States of America*, 63(1):2330, May 1969. ISSN 0027-8424. doi: 10.1073/pnas.63.1.23. URL `http://europepmc.org/articles/PMC534026`.

[223] Laure Fresard and Stephen B. Montgomery. Diagnosing rare diseases after the exome. *Cold Spring Harbor Molecular Case Studies*, 4(6), DEC 2018. ISSN 2373-2873. doi: {10.1101/mcs.a003392}.

[224] Y. A. Barbitoff, D. E. Polev, A. S. Glotov, E. A. Serebryakova, I. V. Shcherbakova, A. M. Kiselev, A. A. Kostareva, et al. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. *Scientific Reports*, 10(1):2057, 2020. doi: 10.1038/s41598-020-59026-y. URL `https://doi.org/10.1038/s41598-020-59026-y`.

[225] M. Spielmann and S. Mundlos. Looking beyond the genes: the role of non-coding variants in human disease. *Human Molecular Genetics*, 25(R2): R157–R165, OCT 1 2016. ISSN 0964-6906. doi: {10.1093/hmg/ddw205}.

[226] C. A. Ibarra Moreno, S. Hu, N. Kraeva, F. Schuster, S. Johannsen, H. Rueffert, et al. An assessment of penetrance and clinical expression of malignant hyperthermia in individuals carrying diagnostic ryanodine receptor 1 gene mutations. *Anesthesiology*, 131(5):983–991, NOV 2019. ISSN 0003-3022. doi: {10.1097/ALN.0000000000002813}.

[227] D. M. Miller, C. Daly, E. M. Aboelsaod, L. Gardner, S. J. Hobson, K. Riasat, S. Shepherd, et al. Genetic epidemiology of malignant hyperthermia

in the UK. *British Journal of Anaesthesia*, 121(4):944–952, OCT 2018. ISSN 0007-0912. doi: {10.1016/j.bja.2018.06.028}.

Appendix

Primer name	Sequence	T_m
***JSRP1* G150A experiment**		
Forward_sequencing	5´ GAA TGA CCA ATT GAA AGA GCG	52.2 °C
Reverse_sequencing	5´ CGA ACT TAG GCT GCA AGA CA	55.1 °C
HRM_1_forward	5´ AGG GAG GGA TGG ACA GTG T	57.9 °C
HRM_3_forward	5´ GGA TGG ACA GTG TGG ACC GT	59.4 °C
HRM_4_forward	5´ ACA GTG TGG ACC GTG CCT C	60.0 °C
HRM_5_forward	5´ TGT CGC TGC AGA CGC CGT	62.9 °C
HRM_1_reverse	5´ CAC ACG TGC TTG GAG TGC T	58.3 °C
HRM_2_reverse	5´ CAC ACG TGC TTG GAG TGC TG	59.2 °C
HRM_3_reverse	5´ GTG CTT GGA GTG CTG CCT CC	61.3 °C
HybProbe_1_forward	5´ GGA TGG ACA GTG TGG ACC	55.3 °C
HybProbe_1_reverse	5´ TCA GTG GAA GGA AGC TCT	52.7 °C
HybProbe_2_reverse	5´ GGT CAG TGG AAG GAA GCT CT	56.6 °C
***PPIP5K1* experiment**		
Forward_sequencing	5´ AAT GGC AGA TGA GAC TGC TTC C	57.3 °C
Reverse_sequencing	5´ GGC TCC TGG ATT CTT CCT CAT AT	56.1 °C
HRM_1_forward	5´ TTG CAG GGG AGG TGA ATG AAG	57.4 °C
HRM_1_reverse	5´ AGC AGC CAG GCC TCA GAT AAC	59.4 °C
***UCP3* sequencing**		
Forward_sequencing	5´ GAC TTC TGC CTA AAT CCC CTT	54.4 °C
Reverse_sequencing	5´ CAT GGC AGT GAA GAC CAG AAT	55.1 °C
***PPARGC1A* sequencing**		
Forward_sequencing	5´ CAC GAA GAC GTG TAT TAT CAA C	51.4 °C

Reverse_sequencing	5´ TGA TGC ATA GTG TTA CCT GCC	54.9 ℃
***RYR1* sequencing**		
Forward_sequencing	5´ CTG TTA CAG AGC AGG TAA GAG	52.3 ℃
Reverse_sequencing	5´ TTC CCA GAT CTC AGG TTT CTG	54.1 ℃
***TLR3* sequencing**		
Forward_sequencing	5´ GTC TCT GAG TAA CAG CCA GC	55.2 ℃
Reverse_sequencing	5´ CGC AAA CTT GTA AAG GAG TTG G	54.6 ℃
***TRPV2* experiment**		
Forward_sequencing	5´ GAG TCA ATG CCT GCA TTC TG	54.1 ℃
Reverse_sequencing	5´ TTG TTA GAA CCC AGG AGT TTG	52.5 ℃
HRM_1_forward	5´ AAT GGG GCC AAT GTG CAT G	56.7 ℃
HRM_1_reverse	5´ CCA AGA AAG GCT CAC TCA C	53.3 ℃
***PRKAA1* experiment**		
Forward_sequencing	5´ CCA TCA ATA CTA CGG AAA TCC	50.6 ℃
Reverse_sequencing	5´ GAC AAG CCC ACC TGA TTC TT	55.1 ℃
HRM_1_forward	5´ CTA TGG GTT TGC CAA ATA TGC	52.2 ℃
HRM_1_reverse	5´ ATG GCA GAA GTA TGT AGA GCA	53.4 ℃

Melting temperatures were calculated using the IDT^{TM} OligoAnalyzer tool

Table A.1: List of primers used in this work

Probe	**Sequence**	**T_m**
Sensor	5´ CGC CGT CCC TCG GGA G -Fluoroscein	65.4 ℃
Anchor	Cy5- 5´ AGC ACT CCA AGC ACG TGT GC -Phosphate	67.6 ℃

Melting temperatures were calculated using the Lightcycler Probe Design Software 2.0 (Roche)

Table A.2: *JSRP1* G150A HybProbe assay probe sequences

Primer set	Forward primer	Reverse primer	Expected product size (bases)
HRM set 1	HRM_3_forward	HRM_2_reverse	100
HRM set 2	HRM_1_forward	HRM_1_reverse	106
HRM set 3	HRM_4_forward	HRM_1_reverse	93
HRM set 4	HRM_5_forward	HRM_1_reverse	49
HRM set 5	HRM_4_forward	HRM_3_reverse	88
HRM set 6	HRM_5_forward	HRM_3_reverse	44
HybProbe set 1	HybProbe_1_forward	HybProbe_1_reverse	175
HybProbe set 2	HRM_1_forward	HybProbe_2_reverse	183

Table A.3: Primer sets for *JSRP1* G150A genotyping

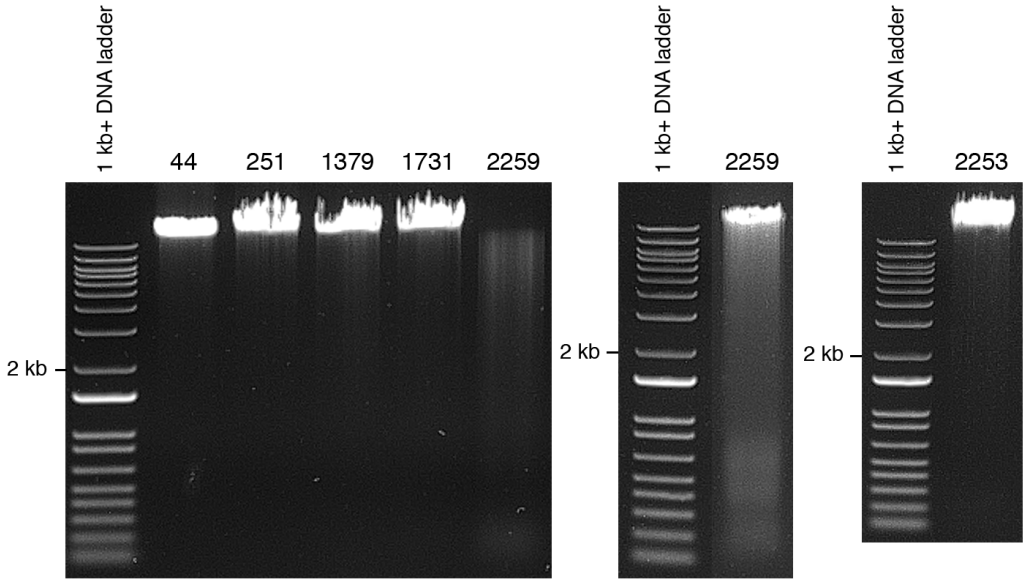

Figure A.1: Quality control of extracted gDNA samples for new WES sequencing of family B and E

250 ng of gDNA from individuals 44, 251, 1379, 1731, 2253 and 2259 were loaded onto a 0.8% agarose gel and the gel was UV imaged to visualise the DNA. The first gel image indicates suitable gDNA quality and quantity for all samples with the exception of sample 2259, with a significant degree of degradation, which was prepared previously from muscle. The second image shows a more recent 2259 gDNA preparation from muscle tissue and indicates improved gDNA quality with less degradation. The third image indicates suitable gDNA quality and quantity for sample 2253.

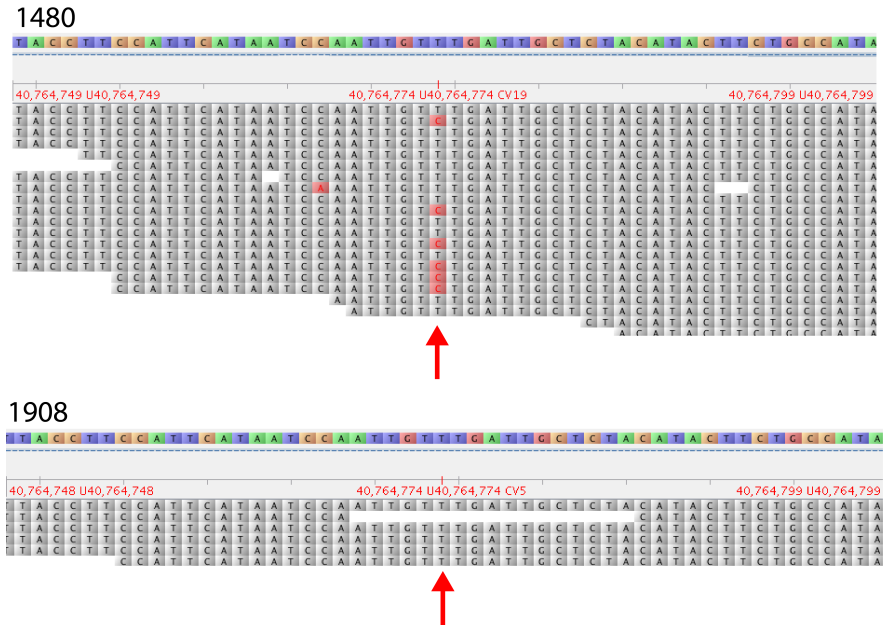

Figure A.2: Sequence alignment of WES data from family A individuals 1480 and 1908 at the genomic position of the *PRKAA1* K429R SNV

The red arrows indicate the variant position. The reads suggest that individual 1480 is heterozygous for the variant, whereas 1908 appears to be wild-type at that position.

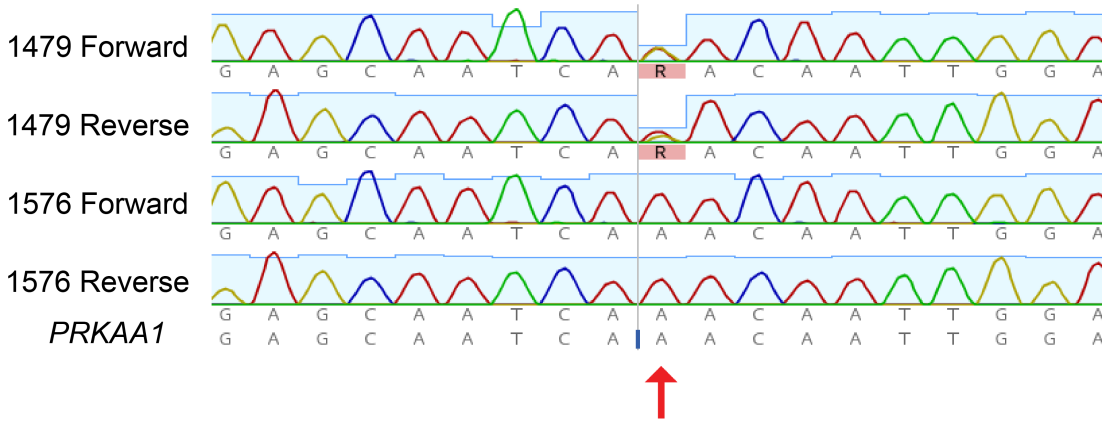

Figure A.3: Alignment of Sanger sequencing results for 1479 and 1576 against the *PRKAA1* nucleotide sequence

The red arrow indicates the variant position. The results indicate that 1479 is heterozygous for the *PRKAA1* K429R variant whereas individual 1576 does not carry variant.

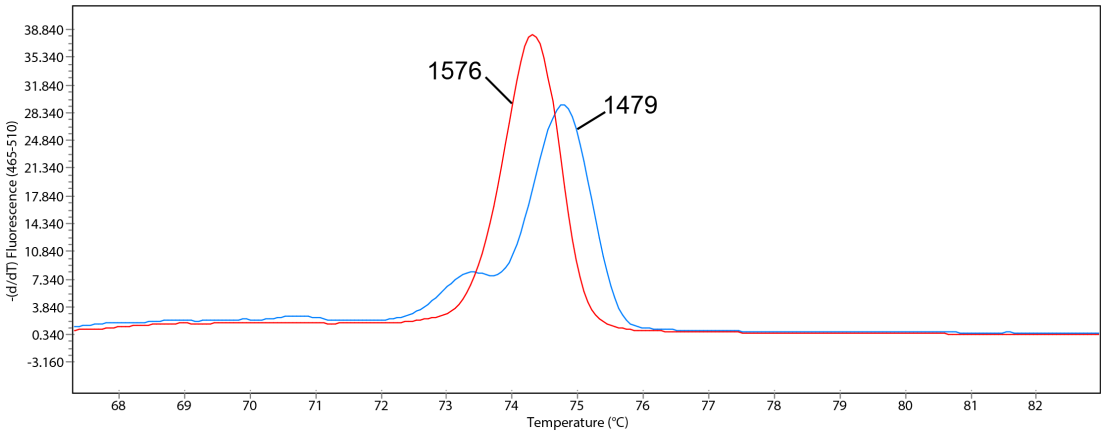

Figure A.4: HRM melting curve for *PRKAA1* K429R variant
Sanger-sequenced wild-type 1576 and heterozygous 1479 have characteristic single and double peaks, respectively.

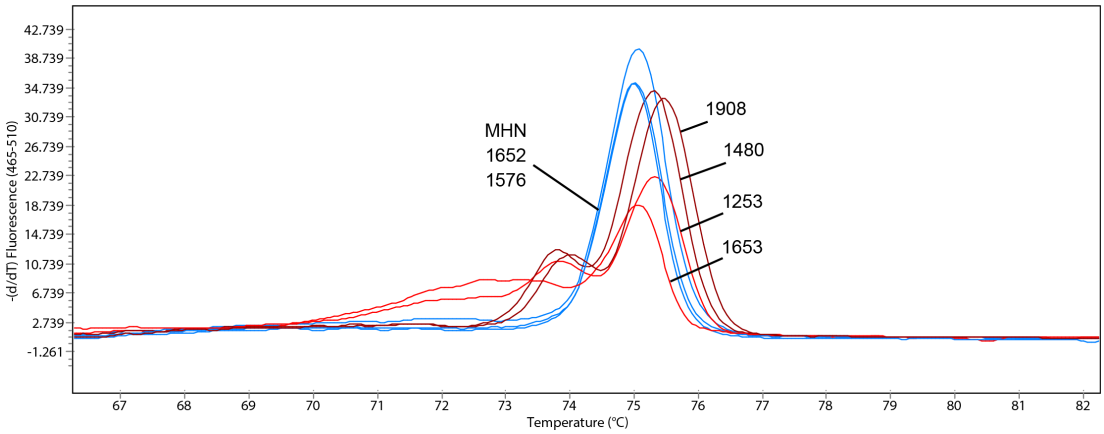

Figure A.5: HRM melting curve for *PRKAA1* K429R variant
The results indicate that individual 1908, thought to be wild-type from the WES data, is heterozygous for the variant.

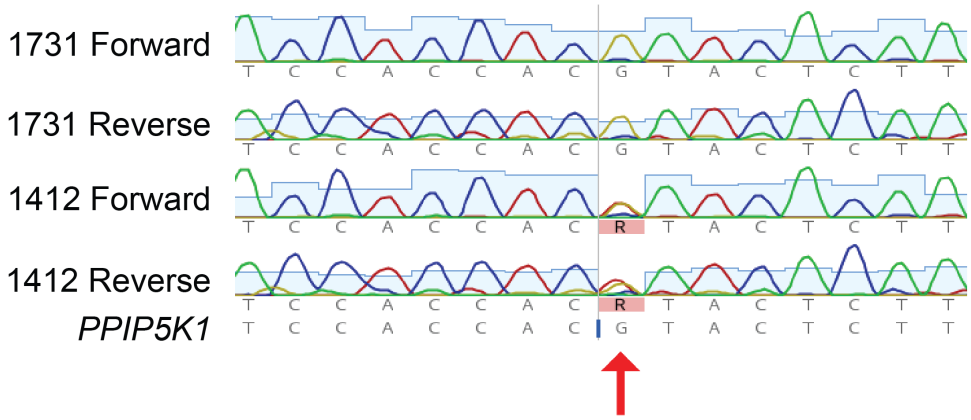

Figure A.6: Alignment of Sanger sequencing results for 1731 and 1412 against the *PPIP5K1* nucleotide sequence

The red arrow indicates the variant position. The results indicate that individual 1731 does not carry the *PPIP5K1* R1081H variant whereas individual 1412 is heterozygous for the variant.

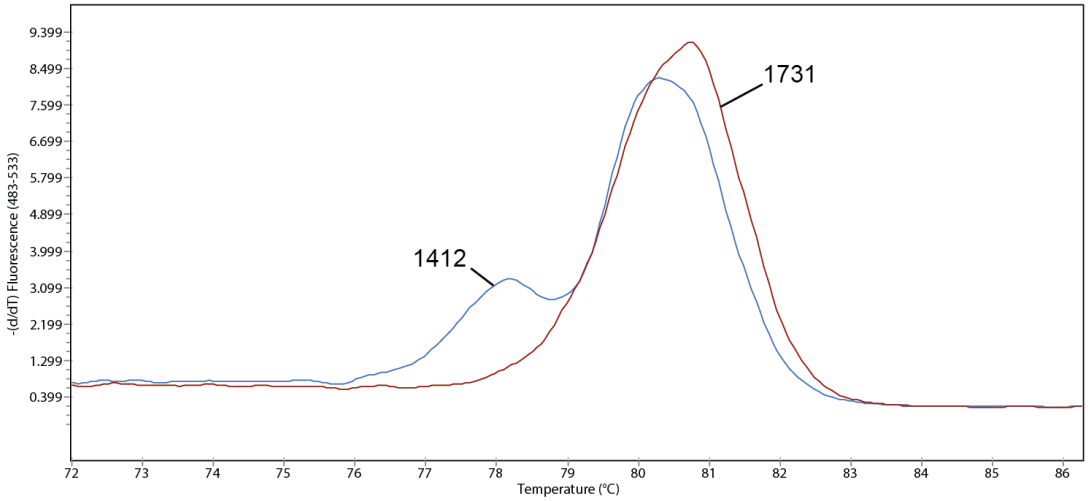

Figure A.7: HRM melting curve for *PPIP5K1* R1081H variant

1731 is confirmed wild-type for the variant and has a characteristic homoduplex single peak. 1412 is confirmed heterozygous for the variant and has a characteristic heteroduplex double peak.

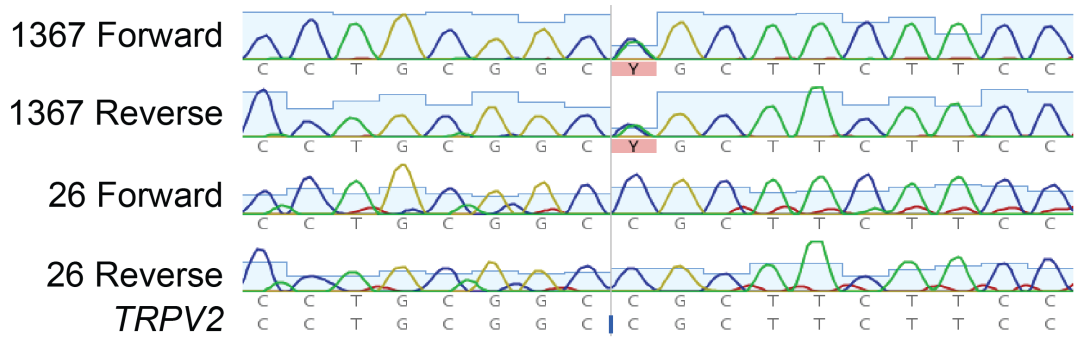

Figure A.8: Alignment of Sanger sequencing results for 1367 and 26 against the *TRPV2* nucleotide sequence

The red arrow indicates the variant position. The results confirm that individual 26 does not carry the *TRPV2* R196C variant whereas individual 1367 is heterozygous for the variant.

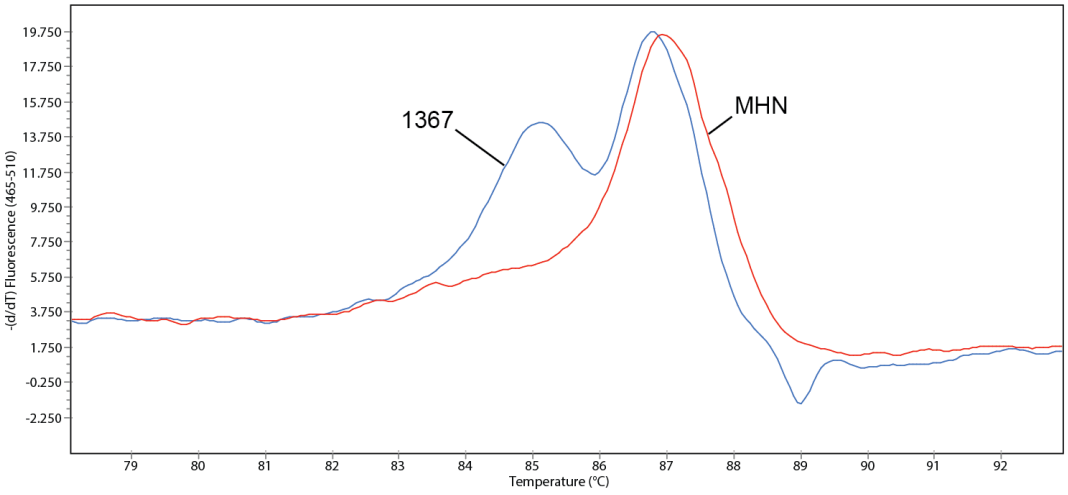

Figure A.9: HRM melting curve for *TRPV2* R196C variant

MHN is wild-type for the variant and has a characteristic homoduplex single peak. 1367 is confirmed heterozygous for the variant and has a characteristic heteroduplex double peak.

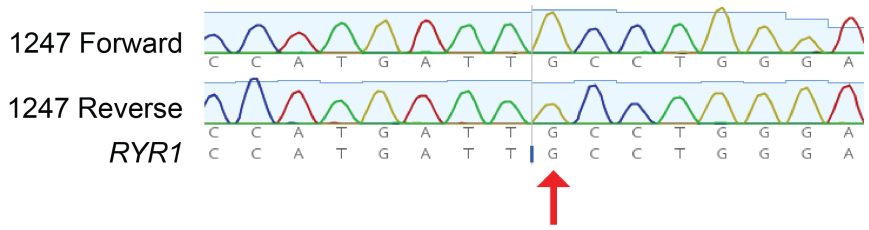

Figure A.10: Alignment of Sanger sequencing results for 1247 against the *RYR1* nucleotide sequence

The red arrow indicates the variant position. The results indicate that 1247 does not carry an *RYR1* E2820Q variant.

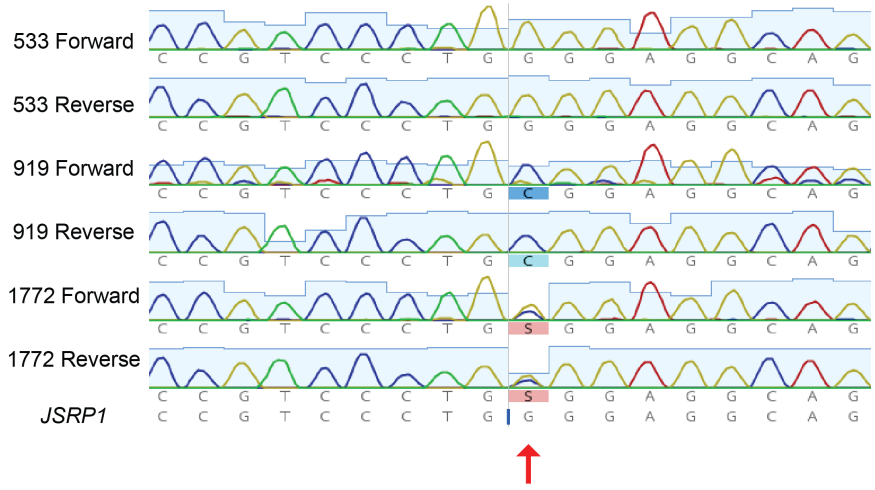

Figure A.11: Alignment of Sanger sequencing results for 533, 919 and 1772 against the *JSRP1* nucleotide sequence

The red arrow indicates the variant position. The results indicate that 533 is wild-type for the variant, 919 is homozygous and 1772 is heterozygous.

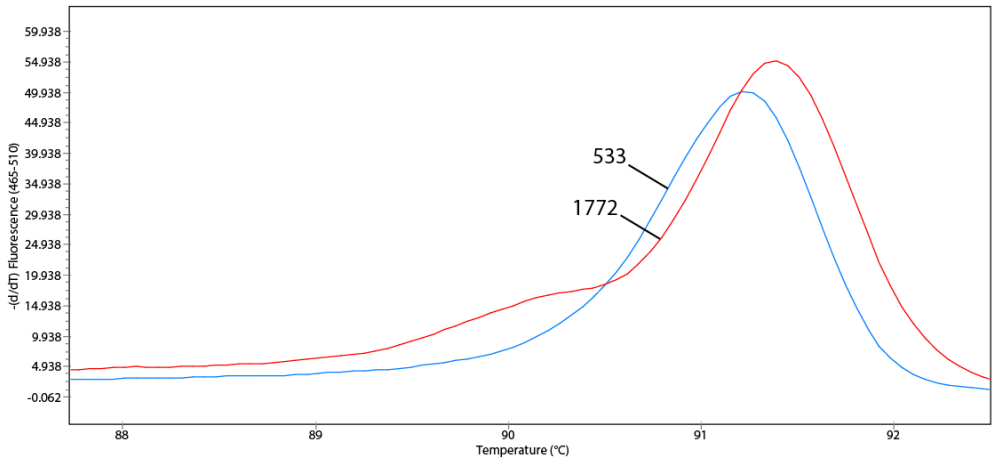

Figure A.12: HRM melting curve for *JSRP1* G150A variant

Wild-type 533 and heterozygous 1772 have single and double peaks characteristic of the presence of homoduplexes and heteroduplexes, respectively.

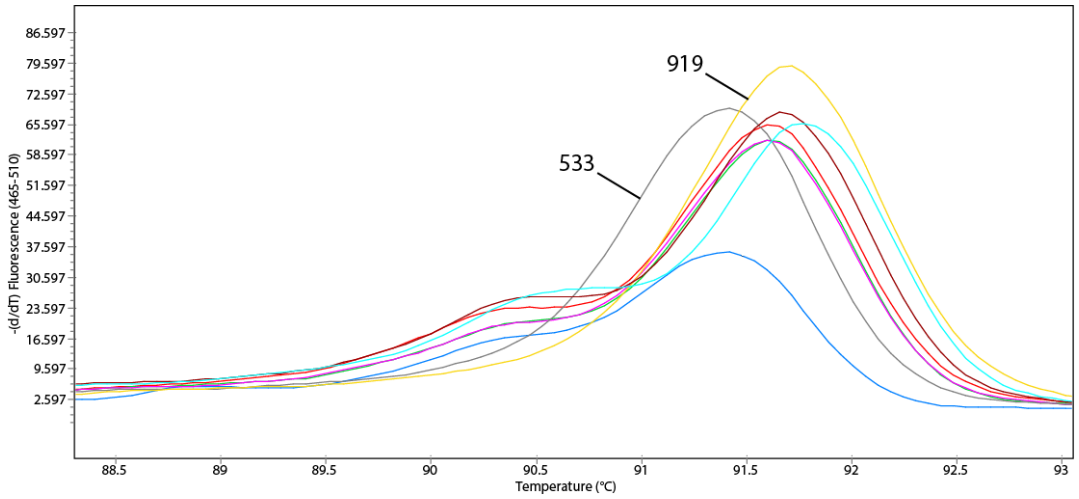

Figure A.13: HRM melting curves for *JSRP1* G150A variant
Wild-type 533 and homozygous 919 have a single peak characteristic of the presence of homoduplexes. Other peaks (not labelled) belong to samples 26, 27, 82, 456, 1367 and 1412.

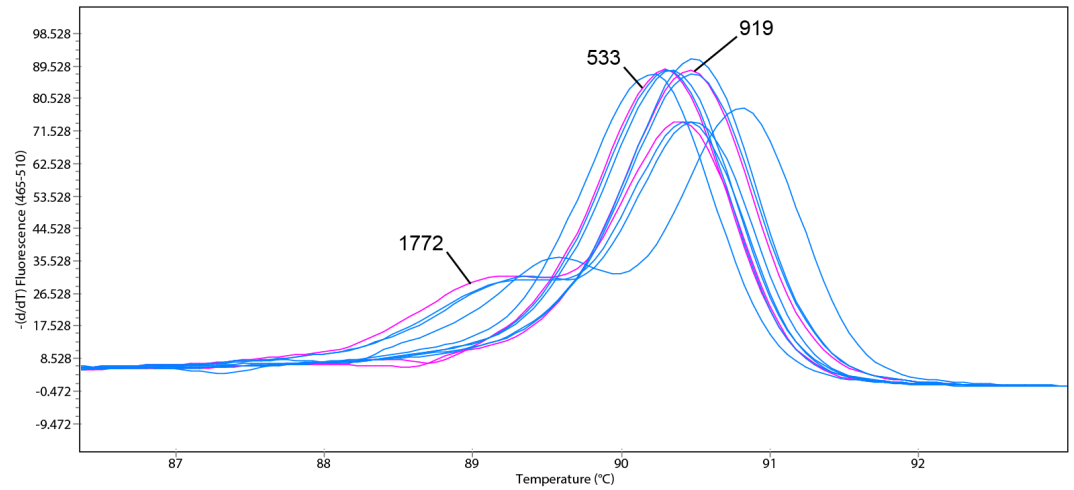

Figure A.14: HRM melting curve for *JSRP1* G150A variant
This assay was carried out with replacement of 0.5 μL H_2O with GC-RICH solution (Roche). Wild-type 533 and homozygous 919 have a single peak characteristic the presence of homoduplexes, and heterozygous 1772 has a double peak characteristic of the presence of a heteroduplex. The shift of homozygous 919 melting peak relative to that of wild-type 533 is no more pronounced than previous assays (see Appendix A.13). Other peaks (not labelled) belong to samples 432, 435, 491, 541, 544, 623, 649 and 685.

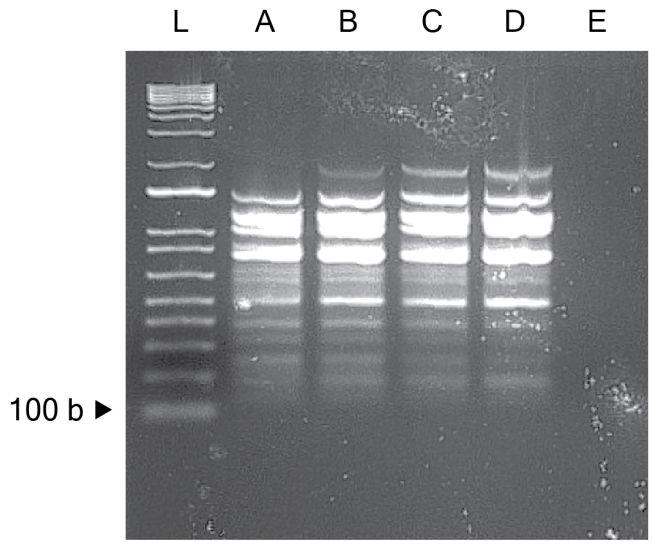

Figure A.15: Agarose gel electrophoresis of PCR products using LightCycler® Probe Design Software 2.0 designed primers for the *JSRP1* G150A HybProbe assay

A, B, C and D are PCRs carried out at annealing temperatures of 52°C, 54°C, 56°C and 58°C, respectively. E is a negative control carried out with the same conditions as B.

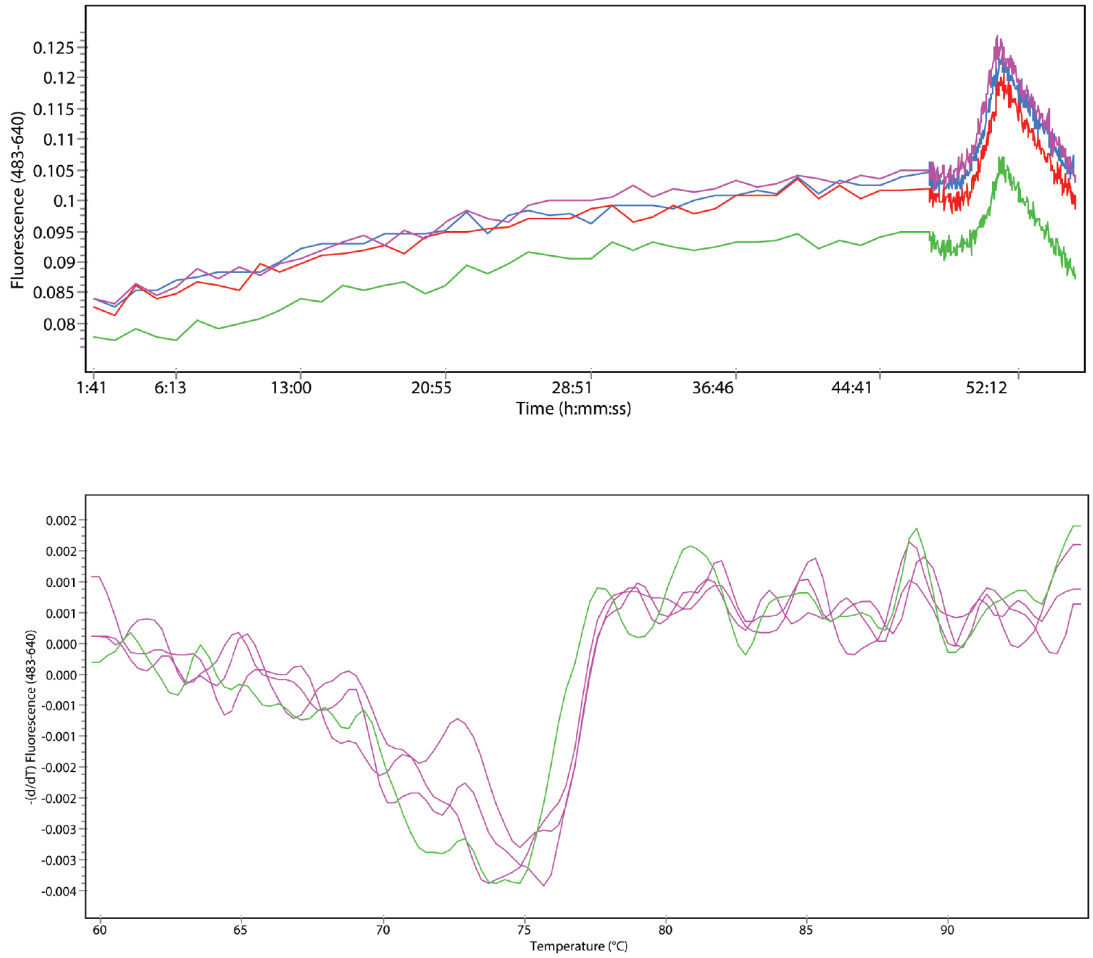

Figure A.16: Representative fluorescence history and melting curve for *JSRP1* G150A HybProbe assays

Samples 533, 919 and 1772, and a negative control, were included in this assay.

		Emission (RFU)			Emission (RFU) (corrected for concentraton)		
	Sample	540 nm	610 nm	700 nm	540 nm	610 nm	700 nm
Blue light excitation (~470 nm)	Diluent (H_2O)	99.45	87.25	85.21	N.A	N.A	N.A
	Sensor (Fluoroscein) 1.25 uM	11652.17	7684.29	1485.17	9.32	6.15	1.19
	Sensor (Fluoroscein) 500 nM	4206.76	2808.78	591.81	8.41	5.62	1.18
	Sensor (Fluoroscein) 250 nM	1606.90	1106.67	278.35	6.43	4.43	1.11
	Sensor (Fluoroscein **new**) 200 nM	3475.28	2282.17	493.23	17.38	11.41	2.47
	Sensor (Fluoroscein **new**) 200 nM	3976.26	2592.36	541.99	19.88	12.96	2.71
Red light excitation (~635nm)	Anchor (Cy5) 100 nM			60068.09			600.68

Figure A.17: HybProbe sensor and anchor probe fluorescence measurements taken on the DeNovix fluorometer

RFU is relative fluorescence units. Emission was corrected for concentration by dividing the RFU values by the probe concenctration in nM to achieve RFU per nM.

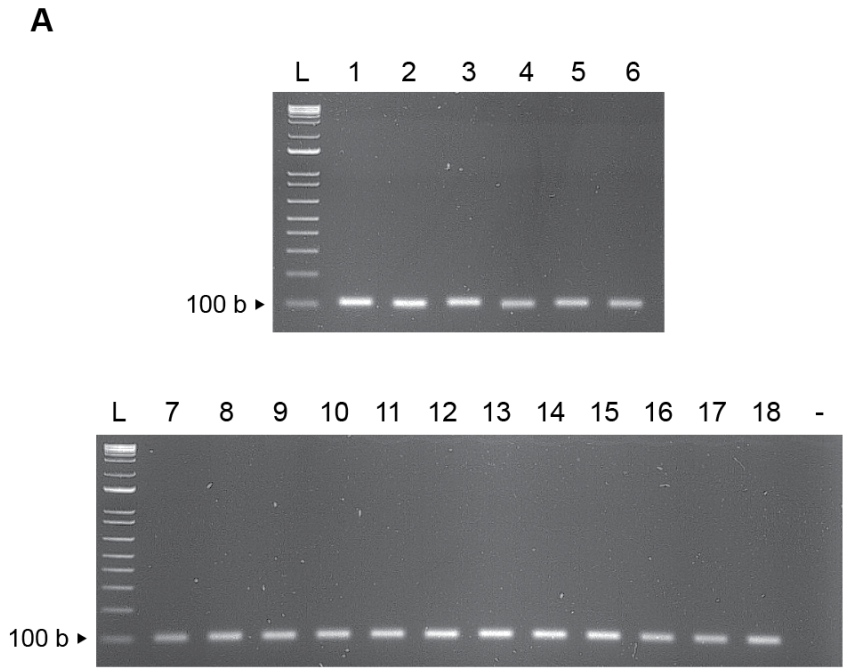

Figure A.18: *JSRP1* G150A RFLP genotyping PCR products set 1
'-' is the negative control (H_2O) lane.

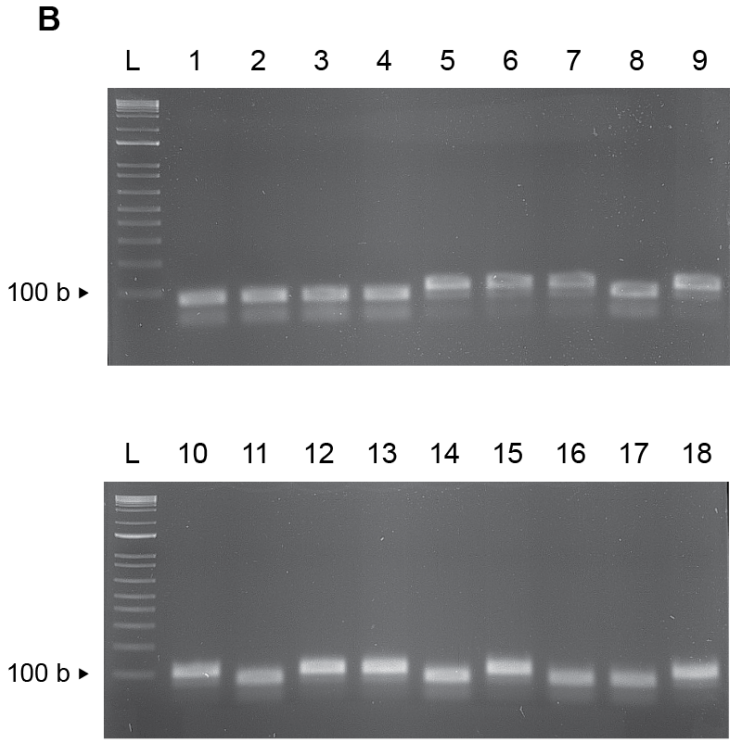

Figure A.19: *JSRP1* G150A RFLP genotyping RFLP products set 1

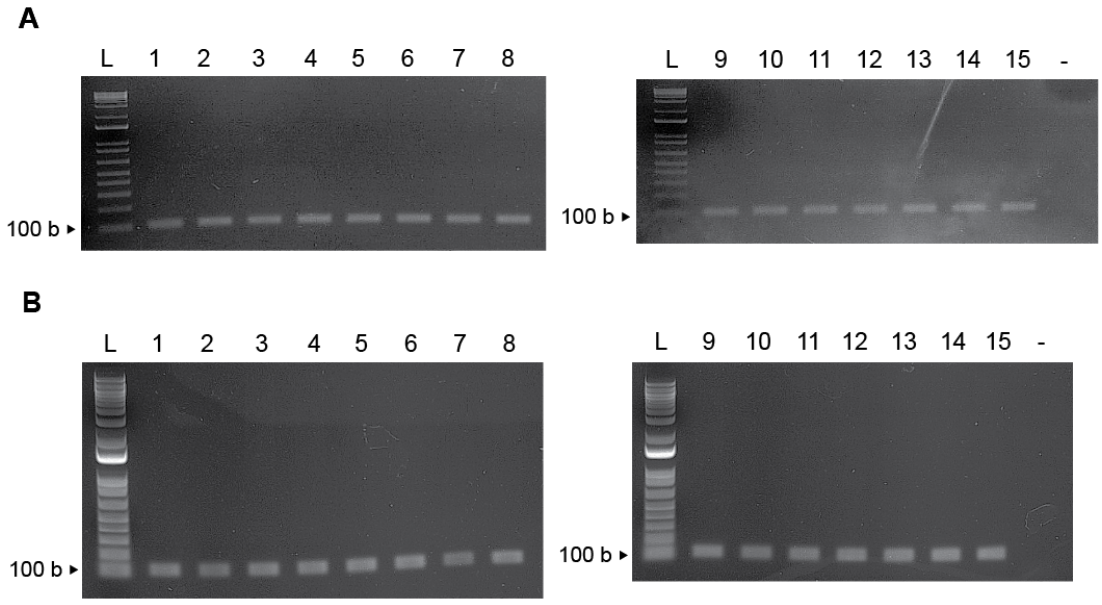

Figure A.20: *JSRP1* G150A RFLP genotyping PCR products set 2
'-' is the negative control (H_2O) lane.

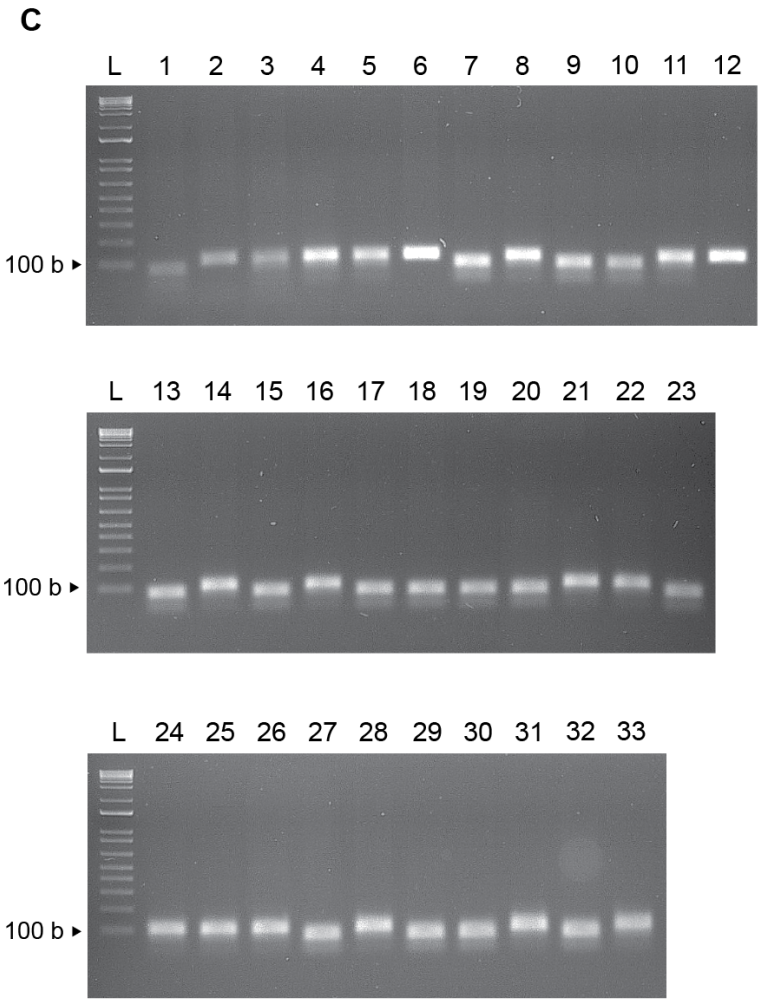

Figure A.21: *JSRP1* G150A RFLP genotyping RFLP products set 2

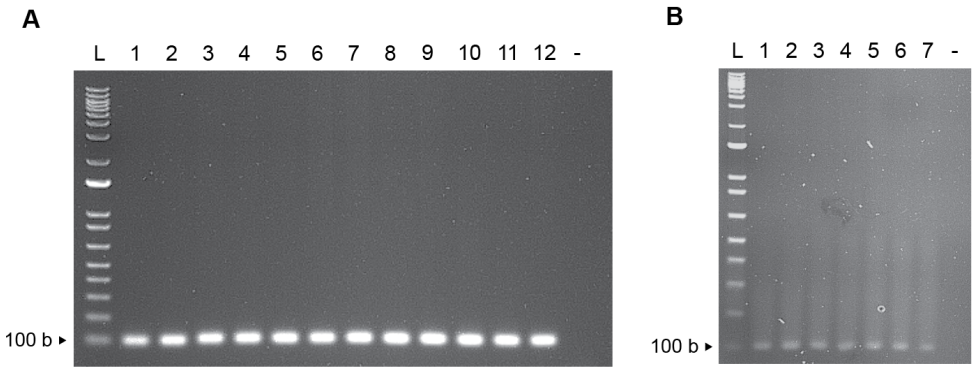

Figure A.22: *JSRP1* G150A RFLP genotyping PCR products set 3
'-' is the negative control (H_2O) lane.

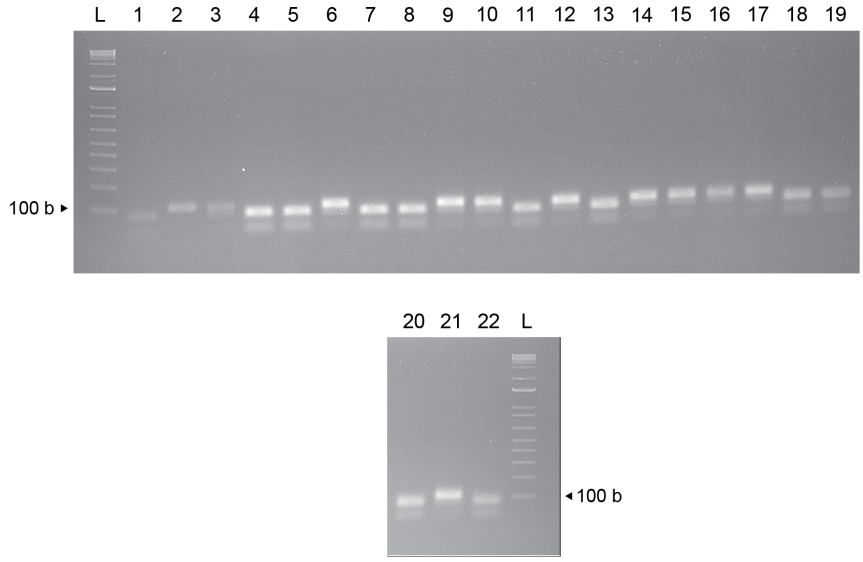

Figure A.23: *JSRP1* G150A RFLP genotyping RFLP products set 3

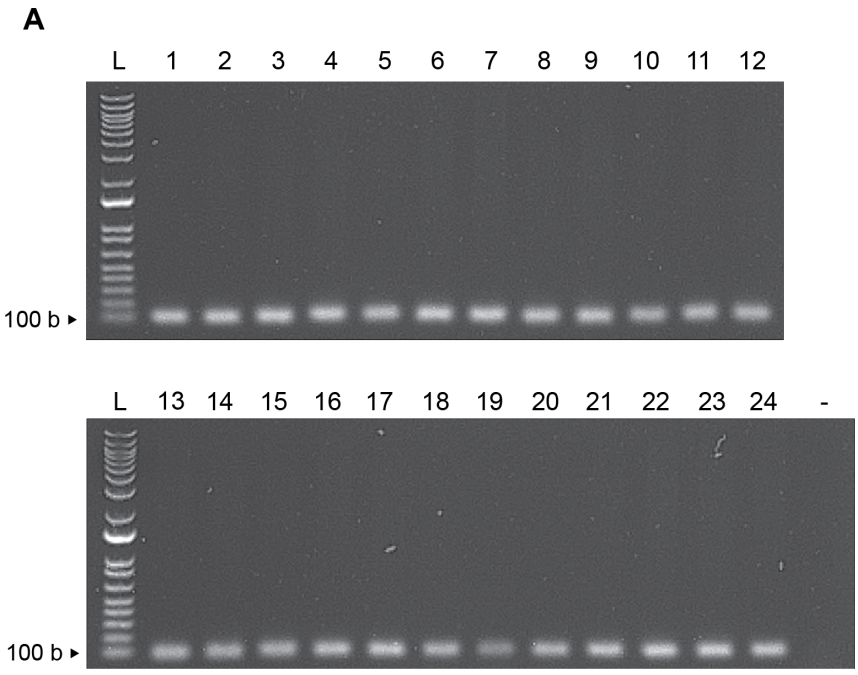

Figure A.24: *JSRP1* G150A RFLP genotyping PCR products set 4
'-' is the negative control (H_2O) lane.

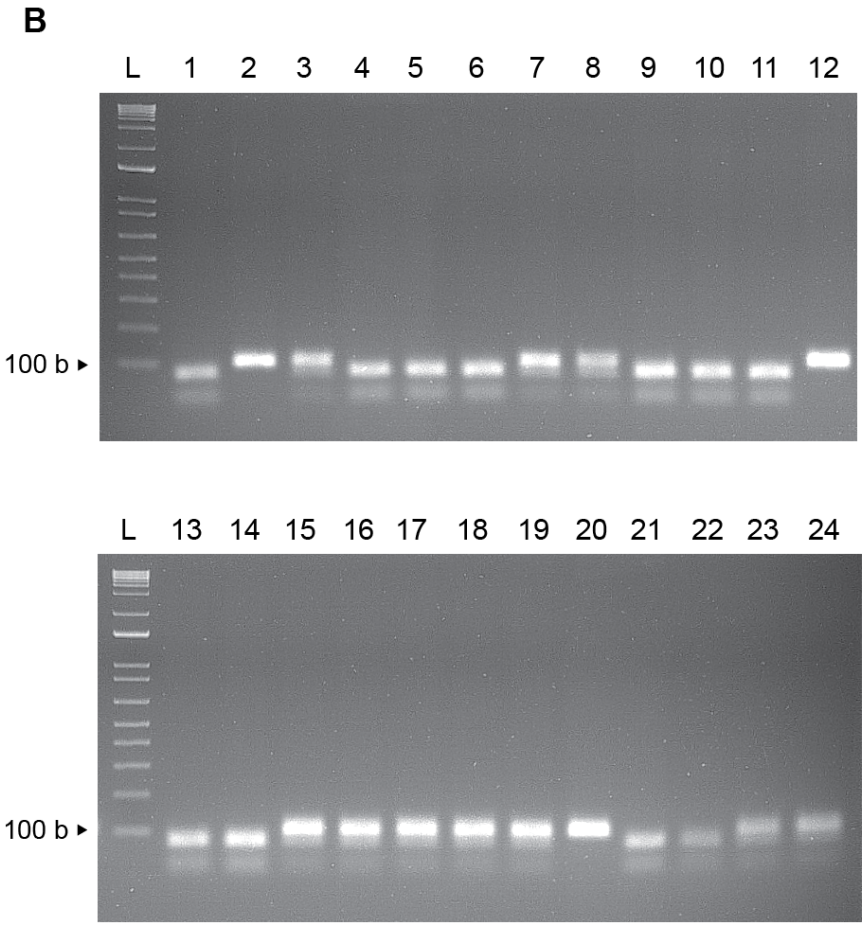

Figure A.25: *JSRP1* G150A RFLP genotyping RFLP products set 4

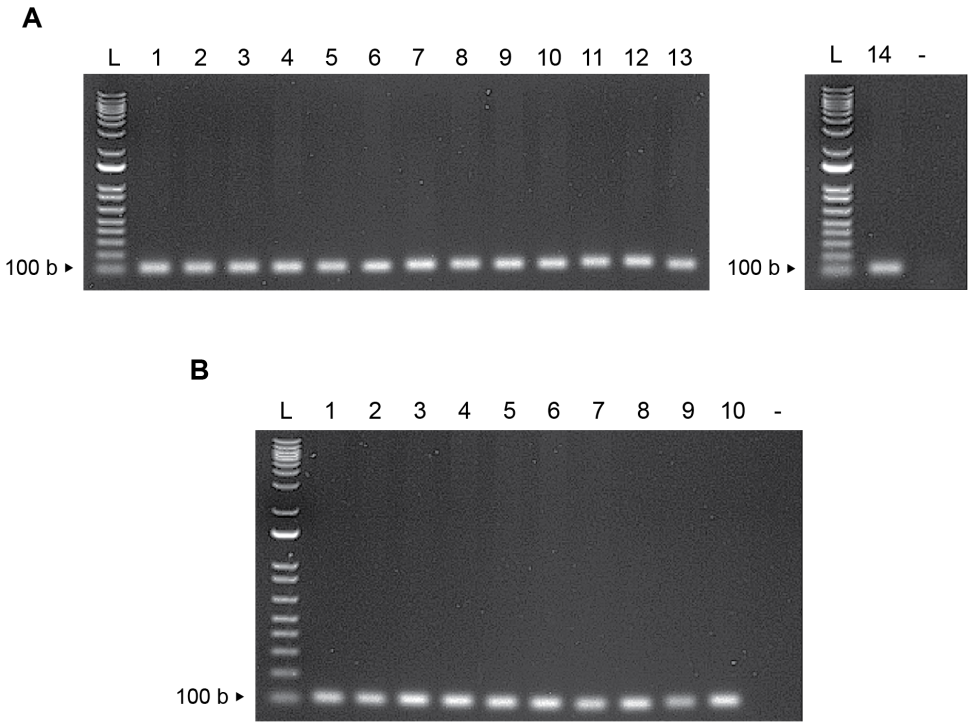

Figure A.26: *JSRP1* G150A RFLP genotyping PCR products set 5
'-' is the negative control (H_2O) lane.

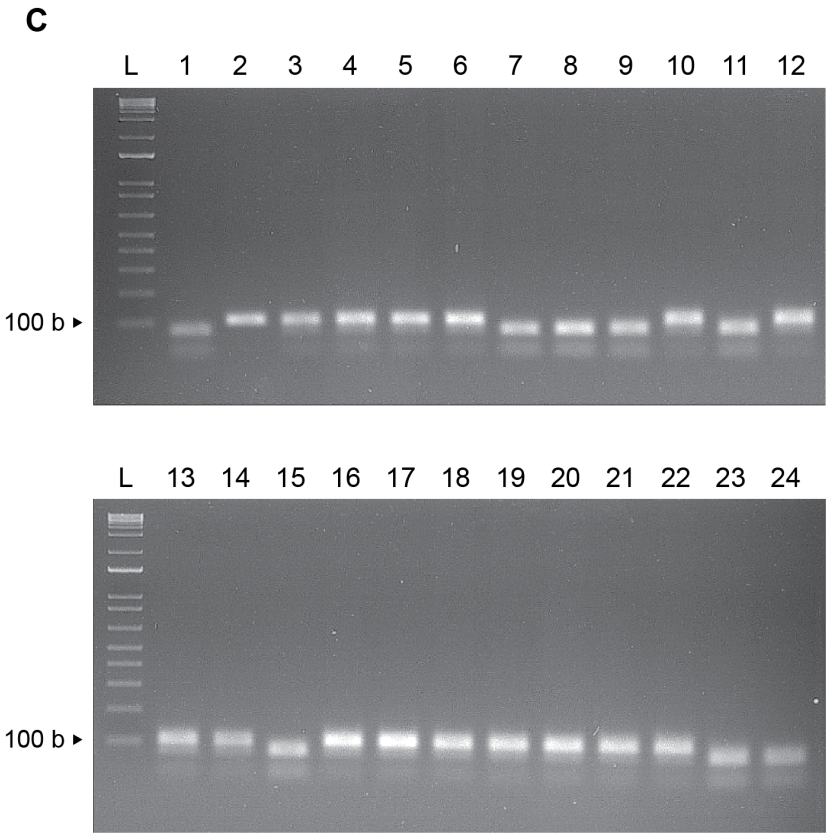

Figure A.27: *JSRP1* G150A RFLP genotyping RFLP products set 5

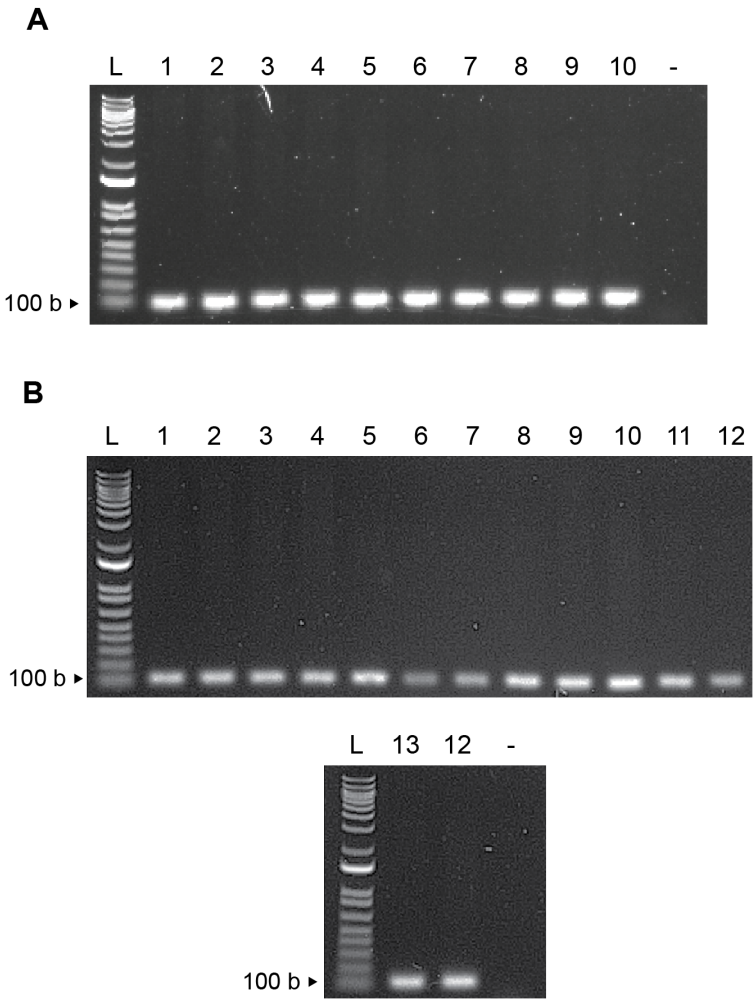

Figure A.28: *JSRP1* G150A RFLP genotyping PCR products set 6
'-' is the negative control (H_2O) lane.

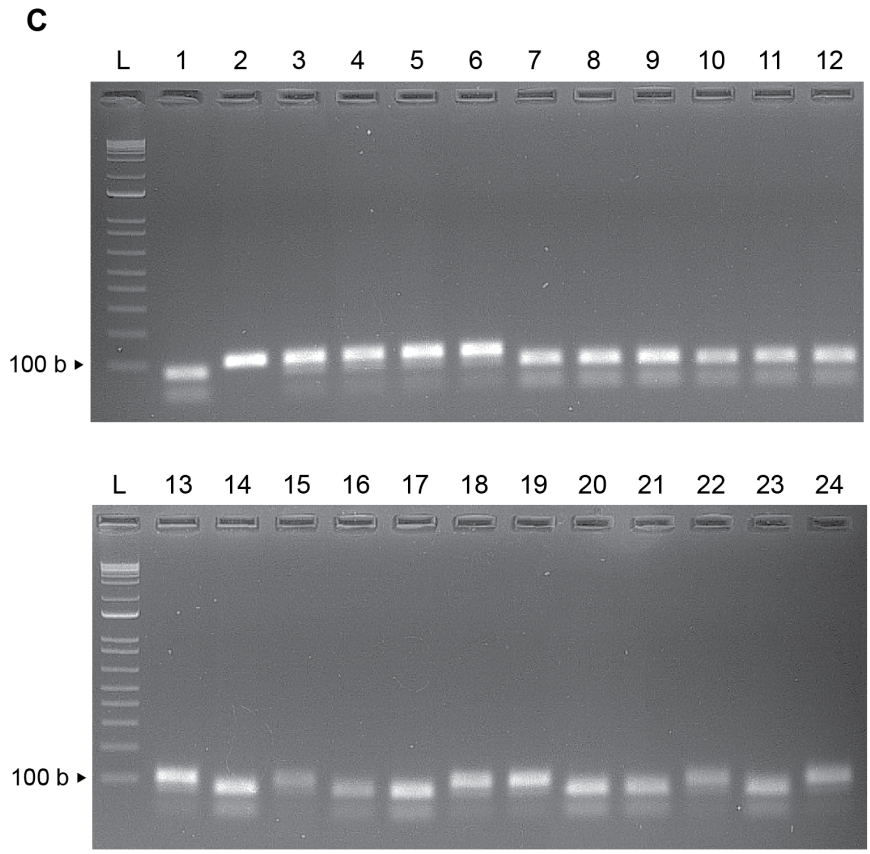

Figure A.29: *JSRP1* G150A RFLP genotyping RFLP products set 6

DNA number	H	C	MH	T4826I	RLFP set	Gel number	G150A
430	0	0.05	MHN	NA	1	1	0
431	3	2.7	MHS	P	1	2	0
432	0.3	0.2	MHN	N	1	3	0
477	0.4	0.6	MHS	N	1	4	0
491	5.3	4.2	MHS	P	1	5	1
494	1.4	1	MHS	N	1	6	1
572	3.8	3.4	MHS	P	1	7	1
623	0.2	0.5	MHE	N	1	8	0
662	0.6	0.25	MHS	N	1	9	1
663	0.1	0	MHN	N	1	10	1
664	0.45	0	MHN	N	1	11	0
685	0.15	0	MHN	N	1	12	1
691	4.6	4.3	MHS	P	1	13	1
1632	0.4	0.2	MHN	N	1	15	1
1637	0	0.04	MHN	N	1	16	0
1731	2.6	2	MHS	N	1	17	0
2	1.6	1	MHS	P	2	4	1
5	0.8	0.8	MHS	P	2	5	1
6	0	0	MHN	N	2	6	2
8	0.2	0	MHN	N	2	7	0
10	6	3.3	MHS	P	2	8	1
22	0.2	0.05	MHN	N	2	9	0
31	0.4	0.2	MHE	N	2	10	0
32	0.4	0.05	MHE	N	2	11	1
39	0.15	0.4	MHE	N	2	12	2
45	3.4	3.4	MHS	P	2	13	0
47	0.3	0.05	MHE	N	2	14	1
50	0.35	0.05	MHE	N	2	15	0
61	5.5	3	MHS	P	2	16	1
86	0.65	0.45	MHS	N	2	17	0
95	1.75	0.75	MHS	N	2	18	0
193	2.4	3.7	MHS	NA	2	19	0
207	0.15	0	MHN	N	2	20	0
1772	1.5	0.5	MHS	N	2	21	1
233	0	0	MHN	N	2	22	1
234	5.9	2.1	MHS	P	2	23	0
239	0	0	NA	P	2	24	1
251	0.2	0	MHN	N	2	25	1
252	0.3	0.5	MHN	N	2	26	1
253	0	0	MHN	N	2	27	0
256	0.6	0.2	MHS	N	2	28	1
274	0.1	0.05	MHN	N	2	29	0
347	5.5	2.5	MHS	P	2	30	0
386	0.4	0.05	MHN	N	2	31	1
349	1	0.25	MHS	N	2	32	0
360	0.3	0	MHN	N	2	33	1
97	0	0	MHN	NA	3	4	0
99	0.65	0.2	MHS	N	3	5	0
113	0.3	0.5	MHN	N	3	6	1
115	0.35	0.15	MHE	N	3	7	0
116	0.3	0.1	MHN	N	3	8	0
118	0.05	0	MHN	N	3	9	1
120	0.05	0	MHN	N	3	10	1
130	1.2	0.3	MHS	N	3	11	0
132	6.6	3.9	MHS	P	3	12	1
143	0.7	0.1	MHE	N	3	13	0
147	6.3	4.5	MHS	P	3	14	1
155	4.5	5	MHS	P	3	15	1
162	0.3	0	MHN	N	3	16	1
175	3.7	0.9	MHS	N	3	17	1
188	5	2.7	MHS	P	3	18	0
189	5.5	3.9	MHS	P	3	19	0
371	1.9	2.5	MHS	P	3	20	0
391	0.2	0	MHN	N	3	21	1

Figure A.30: *JSRP1* G150A association study raw data Table 1

'H' is IVCT halothane contracture value. 'C' is IVCT caffeine contracture value. 'MH' is MH-susceptibility status as determined by IVCT. 'T4826I' is carrier status of the *RYR1* T4826I variant ('P' is heterozygous, 'N' is wild-type and 'NA' not tested). 'RFLP set' the RFLP set number. 'Gel number' is the well number of the RFLP gel. 'G150A' is the *JSRP1* G150A variant carrier status inferred by the RFLP results. Individual 1251 given in this table appears homozygous for the variant by RFLP analysis but was not confirmed by Sanger sequencing and so was not included in the final analysed data.

533	1.7	1.2	MHS	N	3	22	0
713	0.65	0.3	MHS	N	4	4	0
729	0.3	0	MHN	N	4	5	0
751	0	0	MHN	N	4	6	0
828	0.2	0.1	MHN	N	4	7	1
944	0.4	0.2	MHN	N	4	8	1
1007	0.05	0.05	MHN	N	4	9	0
1010	0.4	1.4	MHE	N	4	10	0
1199	0.3	0.2	MHN	N	4	11	0
1251	0.4	0.2	MHN	N	4	12	2
1252	0	0	MHN	N	4	13	0
1282	0.2	0.2	MHN	N	4	14	0
1549	0.4	0.2	MHN	N	4	15	1
1648	0.3	0.1	MHN	N	4	16	1
1681	0.4	0.15	MHE	N	4	17	1
1710	0.2	0.1	MHN	N	4	18	1
1875	0.24	0.04	MHN	N	4	19	1
2075	0.4	0	MHN	NA	4	20	2
2079	0.16	0.12	MHN	N	4	21	0
2081	0.05	0	MHN	N	4	22	0
2188	0.5	0.24	MHS	N	4	23	1
2324	0.28	0.16	MHN	N	4	24	1
27	0.4	0	MHE	N	5	4	1
93	0.7	0.1	MHE	N	5	5	1
121	5.7	4	MHS	P	5	6	1
184	0.1	0.05	MHN	NA	5	7	0
248	0.1	0.1	MHN	N	5	8	0
370	0	0	MHN	N	5	9	0
379	0.02	0	MHN	N	5	10	1
380	0.3	0.05	MHN	N	5	11	0
403	0.4	0	MHE	N	5	12	1
435	0.6	0	MHE	N	5	13	1
469	0.4	0.1	MHN	N	5	14	1
480	0.1	0	MHN	N	5	15	0
490	0.3	0	MHN	N	5	16	1
529	2.6	0.95	MHS	N	5	17	1
601	3.7	1.5	MHS	P	5	18	1
622	0	0	MHN	NA	5	19	1
671	0.9	0.1	MHE	N	5	20	1
675	0	0	MHN	N	5	21	1
978	0.2	0	MHN	NA	5	22	1
1213	0.3	0.2	MHN	N	5	23	0
1242	0.2	0.2	MHN	N	5	24	0
107	0.2	0	MHN	N	6	4	1
266	0.4	0.1	MHN	N	6	5	1
357	3.3	2.8	MHS	P	6	6	1
541	2.1	1.7	MHS	P	6	7	0
554	5.6	3.6	MHS	P	6	8	0
649	0.6	0.2	MHS	N	6	9	0
968	1.3	0.6	MHS	N	6	10	0
1014	1.1	0.7	MHS	N	6	11	0
1110	0	0	MHN	N	6	12	0
1367	2.6	1.7	MHS	N	6	13	1
1417	1.5	0.6	MHS	N	6	14	0
1629	0.28	0.16	MHN	N	6	15	1
1633	0.04	0	MHN	N	6	16	0
1531	0.4	0.1	MHN	N	6	17	0
1675	0.6	0.7	MHE	N	6	18	1
1766	1.4	0.5	MHS	N	6	19	1
1823	0.4	0.2	MHN	N	6	20	0
1975	0.1	0	MHN	N	6	21	0
2064	0	0.04	MHN	N	6	22	1
2101	0.15	0.05	MHN	N	6	23	0
2272	0.08	0	MHN	NA	6	24	1
919	1.4	1.6	MHS	N	NA	2	2

Figure A.31: *JSRP1* G150A association study raw data Table 2

Tool/database	Version/release date
ANNOVAR	2018-04-16
UMD-Predictor	v2
Bedtools	v2.25.0
Bam Util	1.0.14
IGV	2.3.72
HSP API	v1
DbNSFP	3.3a
1000 Genomes	2015-08
DbSNP	147
Clinvar	2019-04-17
Cosmic	v88
GnomAD	v2.11
Iranome DB	2019
Great Middle East Database	2016-10-24
HRCR1	2015-12-03
KAVIAR	2015-09-23
Refgene	2019-04-17
Ensgene	2018-05-06
OMIM	2019-04
HPO	2019-04
Gtex	2018-12
Gene Ontology	GSEA MSigDN 6.2
KEGG	GSEA MSigDN 6.2
Reactome	GSEA MSigDN 6.2
PID	GSEA MSigDN 6.2
Gene Damaging index	2015-10-28
GHIS	from paper
LofTools	from ExAC 0.3
RVIS	v2

Figure A.32: Variant annotation tools and database versions used by VarAFT

Family	Sample	MH status	Halothane	Caffeine	Kit manufacturer/version	Year sequenced
A	1480	MHS	1.8	0.9	SureSelect v6	2017
	1253	MHS	0.6	0.4	SureSelect v6	2016
	1908	MHN	0.3	0.2	SureSelect v6	2016
	1652	MHN	0.3	0.1	SureSelect v6	2016
	1653	MHN	0.3	0.2	SureSelect v6	2016
B	82	MHS	n.d	n.d.	SureSelect v6	2016
	1412	MHS	n.d.	n.d.	SureSelect v6	2016
	456	MHS	n.d.	n.d.	SureSelect v6	2016
	27	MHS(h)	0.4	0	SureSelect v6	2017
	26	UK	n.d.	n.d.	SureSelect v6	2017
	1772	MHS	1.5	0.5	Nimblegen v3	2012
	1367	MHS	2.6	1.7	SureSelect v6	2016
	533	MHS	1.7	1.2	SureSelect	2010
	919	MHS	1.4	1.6	Nimblegen v3	2012
	251	MHN	0.2	0	SureSelect v6	2019
	1731	MHS	2.6	2	Nimblegen v3	2012
	1731_new	MHS	2.6	2	SureSelect v6	2019
C	1360	MHS	1.5	1	SureSelect v6	2017
	1928	MHN	0.3	0	SureSelect v6	2017
	1273	MHS	2.8	1.8	SureSelect v6	2017
	1474	MHN	0.4	0.1	SureSelect v6	2017
	2388	MHS	2.16	0.84	SureSelect v6	2018
D	1128	MHS	n.d.	1.1	Nimblegen v3	2012
	BW1	MHS	1.55	1.1	Nimblegen v3	2012
	932	MHN	0	0	Nimblegen v3	2012
	1172	MHS	4.2	1.5	Nimblegen v3	2012
E	1947	MHS	1.5	0.4	Nimblegen v3	2012
	2253	MHS	2	1.2	SureSelect v6	2019
	1247	MHS	0.6	0.6	SureSelect v6	2019
	1379	MHS	1.2	0.5	Nimblegen v3	2012
	MS1	MHS	0.2	0.3	Nimblegen v3	2012
	44*	MHS	0.98	0.27	SureSelect v6	2019
	1680	MHN	0.2	0	Nimblegen v3	2012
	2259	MHN	0.2	0	SureSelect v6	2019
F	4003*	MHS	.	.	Nimblegen v3	2011
	4004*	MHS	.	.	Nimblegen v3	2011
	4005*	MHS	.	.	Nimblegen v3	2011

Figure A.33: Table of information for all WES samples
SureSelect v6 is the SureSelect^{XT} Human All Exon V6 (Cat. No. 5190-8863, Agilent) kit. Nimblegen v3 is the Nimblegen SeqCap v3 (Cat. No. 06465684001, Roche Nimblegen) kit. Sureselect is the SureSelect Human All Exon (Agilent). * is samples which do not have a lab DNA number.